



HAL
open science

(Co-)Variation of parameters in a Bayesian Situation -An analysis of students' reasoning about the effect of base rate changes

Theresa Büchter, Katharina Böcherer-Linder, Andreas Eichler, Markus Vogel

► To cite this version:

Theresa Büchter, Katharina Böcherer-Linder, Andreas Eichler, Markus Vogel. (Co-)Variation of parameters in a Bayesian Situation -An analysis of students' reasoning about the effect of base rate changes. Twelfth Congress of the European Society for Research in Mathematics Education (CERME12), Feb 2022, Bozen-Bolzano, Italy. hal-03751813

HAL Id: hal-03751813

<https://hal.science/hal-03751813>

Submitted on 15 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

(Co-)Variation of parameters in a Bayesian Situation – An analysis of students’ reasoning about the effect of base rate changes

Theresa Büchter¹, Katharina Böcherer-Linder², Andreas Eichler³ and Markus Vogel⁴

¹University of Kassel, Institute of Mathematics, Germany; tbuechter@mathematik.uni-kassel.de

²University of Freiburg, Institute of Mathematics, Germany; boecherer-linder@math.uni-freiburg.de

³University of Kassel, Institute of Mathematics, Germany; eichler@mathematik.uni-kassel.de

⁴Heidelberg University of Education, Institute of Mathematics, Germany; vogel@ph-heidelberg.de

Evaluating the risk of a hypothesis given some indicators for the hypothesis is a crucial example for conditional probability reasoning. Calculating the probability of a risk when a set of parameters (e. g. the so-called base rate, true- and false-positive rate) is given, is a task which is referred to as a Bayesian task, as it can be solved with the Bayes’ formula. The conceptual understanding of a mathematical formula (and of Bayesian tasks more specifically) implies being able to reason about effects of changes in the given quantities. Based on the dimensions of the concept of functional thinking we propose to refer to this aspect as “Covariation” of Bayesian reasoning. However, hardly any studies have so far investigated Covariation in a Bayesian situation. In this paper we present a study in which participants are asked to reason about changes of the base rate and introduce a coding system with which their answers can be analyzed.

Keywords: Bayesian reasoning, covariation in Bayesian situations, statistics education.

Introduction

Evaluating Bayesian situations, i. e. the risk of a hypothesis (H) given some indicators (I) for the hypothesis, is important for experts from different fields (e. g. medicine and law) in everyday practice and a crucial part of conditional probability reasoning. The comprehension of Bayesian situations has so far almost exclusively been tested with the capability to calculate a probability for the positive predictive value (PPV), that is $P(H|I)$. This is a conditional probability indicating that a hypothesis H (e. g. a medical condition) is true, if an indicator I for the hypothesis (e. g. a positive test result in a medical test) is given. This probability can be calculated with the Bayes’ formula when the base rate of the hypothesis $P(H)$, the true-positive rate $P(I|H)$ and the false-positive rate $P(I|\bar{H})$ of the indicator are given:
$$P(H|I) = \frac{P(H) \cdot P(I|H)}{P(H) \cdot P(I|H) + P(\bar{H}) \cdot P(I|\bar{H})}$$

Summarizing the results of previous works (Binder et al., 2021; Böcherer-Linder et al., 2017; Böcherer-Linder & Eichler, 2017) of our ongoing research program, we argue that three aspects are necessary for a comprehensive understanding of a Bayesian situation: Being able to calculate one specific probability such as $P(H|I)$ in a Bayesian situation using Bayes’ formula (we refer to this aspect as “Performance”), to assess the influence of changes of the given parameters of the situation (we refer to this aspect as “Covariation”) and to discuss the result’s impact (we refer to this aspect as “Communication”). While the aspect of Performance has been studied repeatedly, we are aware of hardly any studies for the remaining two aspects.

As part of a larger project we test and train all aspects of understanding a Bayesian situation (http://bayesianreasoning.de/en/bayes_en.html). In the training, participants learn to systematically combine two beneficial strategies, first representing the statistical information of the Bayesian situation in natural frequencies (McDowell & Jacobs, 2017) and second depicting the structure of the problem in a suitable visualization (e. g. Binder et al., 2020; Böcherer-Linder & Eichler, 2019; Khan et al., 2015). In this paper we present insights from a study which was conducted in preparation for designing these trainings. Thereby, we focus on the aspect of Covariation which we regard referring to the medical situation similar to the one shown in figure 1.

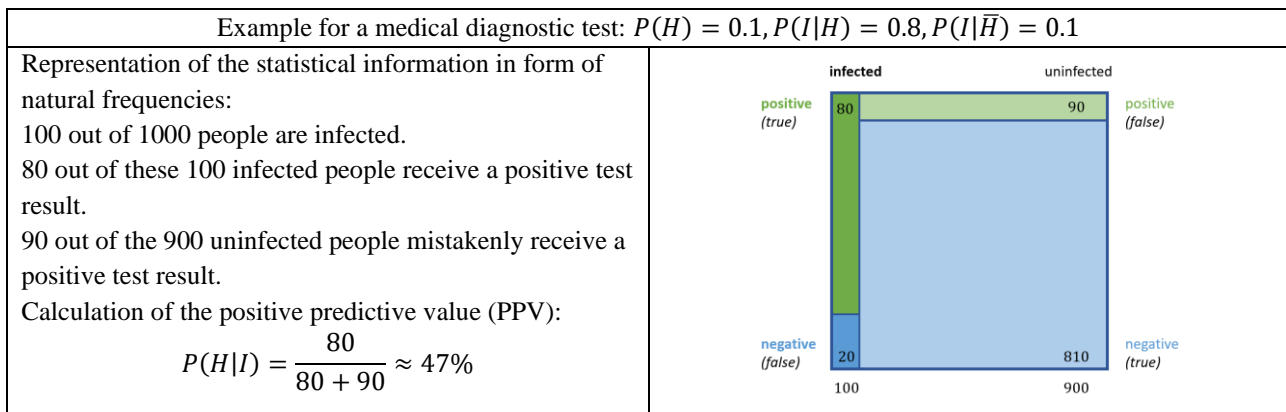


Figure 1: Example for the two beneficial strategies representing the statistical information with natural frequencies and depicting the situation in a visualization, here the unit square

To the best of our knowledge, Covariation as an aspect of Bayesian problems has only been tested once before (Böcherer-Linder et al., 2017). There, it was demonstrated that the effect of base rate change – that is, an increase or a decrease of $P(H)$ - is easier to understand with a unit square than with a frequency tree as a supporting visualization. Yet, this study does not shed light on the question why participants made mistakes and how they approached these questions.

Therefore, we want to provide an insight into first results of our study in which participants’ reasonings for evaluating the change of the base rate in a Bayesian situation are analyzed. Our main question in this paper is: How can students’ reasonings about Covariation (with a focus on a base rate change) in a Bayesian situation be categorized? In order to answer the question, we introduce a category system in this paper with which it is possible to analyze and cluster the level of different reasonings for evaluating the change of the base rate.

Theoretical Background

“It is believed that teaching students how to perceive formulas as covariational entities based on the provided context is essential. This skill can enable them to consider formulas as dynamic functions” (Sokolowski, 2021, p. 184). Even though this quote by Sokolowski has been formulated in the context of teaching physics, we argue that the covariational understanding of a function is equally important in conditional probability reasoning. We thereby refer to Borovcnik (2012, p. 21) who proposes “to investigate the influence of variations of input parameters on the result [i.e. $P(H|I)$]” aiming to strengthen a conceptual understanding instead of a more superficial numerical understanding of the concept of conditional probability and thereby of Bayesian situations.

Hence, we understand the Bayes' theorem not only as a formula but as a function that expresses the dependence of the posterior probability on three parameters. By appreciating Bayesian situations in the context of functions we combine two fields of mathematics education, i.e. statistical education with functional thinking. Consequently, we comprehend the central aspect of functional thinking, i.e. *Covariation* (Lichti & Roth, 2019), also as a part of Bayesian thinking. Covariation stresses the dependence of the independent variable on the dependent variable and the association between changes of both. A typical question for Covariation in the field of Bayesian reasoning is the following: "How does the positive predictive value $P(H|I)$ change when the base rate $P(H)$ increases/decreases?" In the following figure 2, it is illustrated in the unit square how the PPV is affected by an increase/decrease of the base rate.

Base rate	100 out of 1.000 (10%)	500 out of 1.000 (50%)
Unit square	<p>infected uninfected</p> <p>positive (true) 80 90 positive (false)</p> <p>negative (false) 20 810 negative (true)</p> <p>100 900</p>	<p>infected uninfected</p> <p>positive (true) 400 50 positive (false)</p> <p>negative (false) 100 450 negative (true)</p> <p>500 500</p>
PPV	$\frac{80}{80 + 90} \approx 47\%$	$\frac{400}{400 + 50} \approx 89\%$

Figure 2: Dependence of the PPV on the base rate illustrated in the unit square

Considering the effects of changes of the base rate is specifically important, as its influence often causes errors and misunderstandings when calculating the PPV (Kahneman & Tversky, 1982). For instance, in the situation of medical diagnostic tests (Fig. 1), very low base rates can cause a counterintuitive statistical phenomenon since in that case false-positive test results are more likely than true-positive test results and therefore the PPV is considerably low despite good test parameters (i.e. high true- and low false-positive rate). In such situations people tend to ignore the influence of the base rate which is called the "base rate neglect" (Kahneman & Tversky, 1982).

Analysing students reasonings about changes of the base rate can be significant in order to identify their conceptual understanding of the Bayesian situation. The SOLO-taxonomy of Biggs and Collies (1982) proposes a model with which it is possible to cluster the Structure of Observed Learning Outcomes (SOLO) into distinct levels. These levels differ with regard to the amount of (relevant) information which is used and linked to the cue in the students' arguments. Thereby, they distinguish five different levels of observed learning outcomes which can be applied to the teaching of various topics. We propose to apply the SOLO taxonomy by Biggs and Collis (1982) to the categorization of reasonings which are given about changes of the base rate in a Bayesian situation. In order to apply this model to tasks on Covariation, we identify what information is relevant for this task and thereby adapt the levels by Biggs and Collies to Covariation tasks in a Bayesian situation.

The PPV $P(H|I)$ in a Bayesian situation is calculated by $P(H|I) = \frac{P(I|H) \cdot P(H)}{P(I|H) \cdot P(H) + P(I|\bar{H}) \cdot P(\bar{H})}$ (Fig. 1).

Thus, when reasoning about changes of the PPV it is necessary to consider alterations in both

quantities representing the multiplied probabilities $P(I|H) \cdot P(H) = P(H \cap I)$ and $P(I|\bar{H}) \cdot P(\bar{H}) = P(\bar{H} \cap I)$ and then analyze their effect on the fraction (i.e. the PPV). Both quantities are dependent on the base rate. Illustrated in the unit squares in Fig. 2, one can see that the amount of infected people with a positive test result represented by $P(H \cap I)$ in the Bayesian formula increases with a higher base rate, whereas the amount of uninfected people with a positive test result represented by $P(\bar{H} \cap I)$ decreases when true- and false-positive rates stay the same. The relative increment of infected people with a positive test result (nominator) is higher than the relative increment of all people with a positive test result (denominator). Therefore, the PPV increases with an increase of the base rate. The different levels in the SOLO-model adapted to a Covariation task are described in figure 3.

Level	Description by Biggs & Collis	Covariation task in a Bayesian problem
Pre-structural	An irrelevant feature might be linked to the cue.	No explanation is given for the described relation between base rate and PPV or irrelevant consequences of the base rate change are described (e. g. on $P(\bar{H} \cap \bar{I})$).
Uni-structural	One relevant feature is linked to the cue.	The effect on only one of the relevant quantities ($P(H \cap I)$ or $P(\bar{H} \cap I)$) is considered or the effect of several quantities in the Bayesian situation is considered, but only one of them is relevant.
Multi-structural	Several relevant features are linked to the cue.	Both relevant quantities ($P(H \cap I)$ and $P(\bar{H} \cap I)$) are considered, but their relation to the PPV is not clearly spelled out.
Relational	All relevant data are considered and put into a conceptual scheme.	Both relevant quantities ($P(H \cap I)$ and $P(\bar{H} \cap I)$) are considered and their relation to the PPV is clearly spelled out.
Extended abstract	All relevant data are considered and subsumed in an abstract model.	A general model for the dependence of the PPV on changes of the base rate is generated.

Figure 3: SOLO model (Biggs & Collis, 1982) applied to Covariation tasks in Bayesian situations

In this study, we want to describe how we developed a coding system for the reasonings about Covariational tasks in Bayesian situations in order to classify them within these categories. Moreover, we will describe which further categories we inductively derived in order to code differences. Consequently, we will outline in the results section how the different levels of reasoning are distributed among preservice teachers, who were the subjects of our study.

Material and methods

Study design

Every participant answered five questions about a Bayesian situation such as the medical situation shown in Fig. 1. Thereby, we used two different Bayesian situations: one context about breathalyzers and another one about a mammography screening. The first task was always to a) calculate the PPV. The consecutive tasks were to determine how b) an increase of the true-positive rate, c) an increase of the false-positive rate, d) a decrease of the base rate and e) an equally large increase of the true- and the false-positive rate simultaneously affected the PPV. Answers were given in form of single choice questions with three options: the PPV i) decreases, ii) stays the same, iii) increases. Task a) was always the first one and task e) was always the last one. Tasks b) to d) were presented in a random order. Each participant was asked to give a reasoning for their choice of answer in one of the single-

choice questions of b) to e). The answer to which a participant was asked to formulate a reasoning was chosen randomly. In the results section we only refer to the reasonings which participants have given for their decision of the effect of the base rate change.

Materials

Since visualizations and natural frequencies help to understand the influence of the base rate (Böcherer-Linder et al. 2017, also compare Fig. 2), we investigate how people reason about Covariation with the help of these two beneficial strategies (Fig. 1). As visualizations, we used a double-tree and a unit square which were more supportive in tasks of Performance than the simple tree diagram (Böcherer-Linder & Eichler, 2019) as we suppose that Performance is a prerequisite for Covariational tasks. Thus, the Bayesian situation was described with probabilities and also displayed in a visualization (unit square or double-tree) with frequencies.

Participants

230 pre-service teachers (181 females, 47 males, 2 unknown) participated in this study. They all study to become teachers in mathematics and another subject but for different age groups (e. g. some for primary school others for secondary school). They did not receive any prior training in stochastics. 60 out of the 230 participants were asked to give a reasoning for their answer to the single choice question about changes of the base rate.

Results

Developing categories for the data analysis

In this section we illustrate, how we coded the different reasonings and their belonging to the different levels of the SOLO taxonomy. Thereby, we refer to four exemplarily reasonings below. The initial question was: “Imagine: The probability that a driver is under the influence of alcohol is actually smaller than 10%. How does that affect the probability that a driver is actually under the influence of alcohol, if (s)he receives a positive test result in the breathalyzer?” After selecting if $P(H|I)$ increases, decreases or remains constant, the participants were asked to explain their choice:

Example 1: “The precision of the test is not changed by the description in the text.”

Example 2: “ $0.1 \cdot 0.9 = 9\%$, something smaller than $0.1 \cdot 0.9 = something smaller than 9\%$ ”

Example 3: “As the number of people who are under the influence of alcohol decreases, the probability also decreases, that a positively tested person is also under the influence of alcohol. The nominator of the fraction decreases and the denominator stays the same. Thus, the result is smaller.”

Example 4: “Number of people under the influence of alcohol decreases, analogously 90% positive \rightarrow less under the influence of alcohol and positive and number of false positively tested bigger. Denominator bigger and nominator smaller therefore result is smaller.”

First, we coded to which probabilities/quantities of the Bayesian situation the reasonings made a reference to. Example 1 refers to the true- and false-positive rate of the test (“precision of the test”), whereas examples 2 and 3 both refer to the quantity of true-positives (the numbers of the Bayesian context were chosen in a distinct way so we made this inference in example 2). In example 4

references to the quantity of true-positives and false-positives are drawn. However, in example 4 we can also observe that the conclusion for the denominator of the fraction for the PPV is wrong (i.e. it actually decreases as well). In total we observed six types of quantity-references (QR), some including further subtypes: no reference to any probability or quantity (e. g. “somehow seems logic”, QR0), references to one or more probabilities of the test (e. g. true- or false-positive rate, QR11-QR14), references to one of the joint probabilities (e. g. true- or false-positives, QR21-QR22), references to two joint probabilities or joint events with only one of them being relevant (QR31-QR32) or both of them being relevant (QR33-QR34), description of a direct link between a decreased base rate and a decrease PPV without further explanation (QR40) and between a decreased base rate and an increased PPV (QR50). The types QR33 and QR34 differ only in their implications on the PPV: while both correctly describe the changes in the two joint probabilities, only QR34 draws the correct conclusions from it while QR33 doesn’t. In figure 4 we display how we have assigned the different observed probability/quantity references to the levels in the SOLO taxonomy.

Level	Observed probability/quantity reference
No level	QR0 and QR50
Pre-structural	QR11-QR14 + QR40
Uni-structural	QR21-QR22 + QR31-QR32
Multi-structural	QR33
Relational	QR34
Extended abstract	No observations.

Figure 4: Observations of quantity/probability references in the different levels of the SOLO model

Apart from differences in quantity references we noted further differences and inductively generated additional categories to quantify these differences. First, we noticed a difference in the reference to the context of the Bayesian situation. While in example 2 there is no reference to the context at all, examples 3 and 4 refer to the specific context of the Bayesian situation. Example 1 refers to a more general context, as the “precision of the test” is just as suitable for the breathalyzer context as for the mammography context. Therefore, we differentiated between the connection of a rationale with a specific context with three codes: no context (C0), specific context (C1) and general context (C2). Second, we observed a difference in the representation of the described (changed or unchanged) quantities/probabilities. In example 2 only percentages are used, in example 4 percentages, references to a fraction and absolute frequencies are used. In example 3 probabilities, references to a fraction and absolute frequencies are used. For each reasoning we coded if each of the following types of representations was used (code 1) or not (code 0): probability (R1), percentage (R2), frequencies (R3), fraction (R4), proportion (R5), quota (R6) and size of an area in the unit square (R7).

Reporting the data within the categories

Two raters independently coded all 60 reasonings for effects of the base rate change with the coding system described above. We report inter-rater-reliability that was assessed for all categories described in a first attempt without further training. For the reference to the context Cohen’s Kappa was 0.75 (95% CI: [0.59;0.9]) and therefore substantial. For each representation an interclass correlation coefficient (ICC) was calculated. The ICC for a quota as a representation was poor with an ICC of 0.381 (95% CI: [0.144;0.577]). Yet, the other representation types were coded with a good to excellent

inter-rater-reliability with a range from 0.687 (95% CI: [0.528;0.8]) for frequency, to 0.92 (95% CI: [0.841;0.94] for proportions. For the quantity-representation the Cohen's Kappa was 0.76 (95% CI: [0.63-0.88]). Overall, apart from the coding of the quotas as representation the inter-rater reliability seems satisfying. The results of the ratings of both raters are represented in figure 5.

	Context			Representation						
	no (C0)	specific (C1)	general (C2)	prob. (R1)	perc. (R2)	frequ. (R3)	frac. (R4)	prop. (R5)	quota (R6)	area (R7)
No level (25/20)	(12/11)	(11/9)	(2/0)	(13/8)	(6/5)	(6/1)	(4/2)	(2/0)	(3/1)	(2/0)
Prestructural (15/23)	(2/9)	(10/11)	(3/3)	(11/13)	(8/8)	(3/2)	(0/2)	(1/4)	(1/0)	(1/0)
Unistructural (18/15)	(1/1)	(15/14)	(2/0)	(10/9)	(10/4)	(9/11)	(3/1)	(2/0)	(0/0)	(1/1)
Multistructural (1/1)	(0/0)	(1/1)	(0/0)	(0/0)	(1/1)	(1/1)	(1/1)	(0/0)	(0/0)	(0/0)
Relational (1/1)	(0/0)	(1/1)	(0/0)	(1/1)	(0/0)	(1/1)	(0/0)	(1/1)	(0/0)	(0/0)

Figure 5: Number of Answers in each of the categories (rater 1/rater 2)

Discussion

Bayesian tasks have so far almost exclusively been studied as tasks of Performance. These tasks are known to be difficult and one of the assumed reasons for that is the base rate neglect, thus the cognitive error by which the influence of the base rate on the PPV is overlooked. We have introduced the concept of Covariation (which is an established dimension of functional thinking) to Bayesian tasks. With Covariation tasks in a Bayesian situation one can directly assess the participants' ability to judge the influence of the base rate on the PPV. Asking for a reasoning of the given answer (as we did in this study) allows to qualitatively study the participants understanding of the Bayesian situation. The coding system which we have introduced in this paper can be of special help when teaching about Bayesian tasks, since arguments which belong to the different levels of the SOLO-model reveal different issues about the understanding of the Bayesian situation. For instance, students who reason according to the prestructural level clearly lack an understanding of the Bayesian situation itself as they are unaware of the relevant quantities which have to be considered in the particular Bayesian task. Therefore, they should revise the basics again (e. g. what sets and subsets are described in the situation and how can they be quantified and used to calculate the PPV). On the other hand, students whose reasoning belongs to the multistructural level are very well aware of the structure of a Bayesian situation. They might only need some support in how to argue about changes in a fraction. Thus, the coding system can help to tailor the support to the students' needs. Moreover, it is evident, that the reasoning level in Bayesian situations is generally rather poor with 40 out of 60 students whose reasoning remains on the pre-structural level or cannot be assigned to either level. This observation confirms prior research about Bayesian reasoning in so far as Bayesian tasks have generally been shown to be considerably challenging and counter intuitive without any training.

A consecutive cluster analysis with the data derived from the coding system together with additional information from the study will reveal if and how the levels in the reasoning about Covariation coincide with other aspects (e. g. the capability to correctly calculate the PPV or the type of

visualization which was used as a supportive tool). This will shed more light on how to successfully teach a conceptual understanding of conditional probabilities.

References

- Biggs, J. B., & Collis, K. F. (Eds.). (1982). *Educational psychology. Evaluating the quality of learning: The SOLO taxonomy; structure of the observed learning outcome*. Acad. Pr. <http://www.sciencedirect.com/science/book/9780120975525>
- Binder, K., Krauss, S., Schmidmaier, R., & Braun, L. T. (2021). Natural frequency trees improve diagnostic efficiency in Bayesian reasoning. *Advances in Health Sciences Education*, 26(3), 847–863. <https://doi.org/10.1007/s10459-020-10025-8>
- 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.v1i11.169>
- 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>
- 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>
- 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>
- 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>
- 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>
- Sokolowski, A. (2021). Are Physics Formulas Aiding Covariational Reasoning? Students' Perspective. In A. Sokolowski (Ed.), *Understanding physics using mathematical reasoning: A modeling approach for practitioners and researchers* (pp. 177–186). Springer. https://doi.org/10.1007/978-3-030-80205-9_12