"Take the Middle" – Averaging Prior and Evidence as Effective Heuristic in Bayesian Reasoning

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

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

Abstract

When humans revise their assumptions based on evidence, they process information on the (un)certainties of the situation. This process can be modeled by a (mathematically optimal) Bayesian reasoning strategy. Humans typically deviate from this norm and apply heuristic strategies, often by only partially processing the available information (e.g., neglecting base rates). From a perspective of ecological rationality, such heuristics possibly constitute viable cognitive strategies in certain situations. We investigate the adequacy of a cognitively plausible heuristic strategy, which amounts to approximately averaging the probability information on prior hypotheses and evidence. We compare this strategy to optimal Bayesian reasoning and to information-neglecting strategies by exploring the situational parameter space (number of hypotheses, prior and likelihood values). Finally, we frame this in the context of teachers' diagnostic judgments on students' potential misconceptions (priors) based on students' solutions (evidence) and interpret the resulting accuracy of decisions within the ecology of informal student assessment.

Keywords: Bayesian reasoning; averaging-prior-and-evidence strategy; diagnostic judgments; ecological rationality

Introduction

When inferences in situations with uncertainty are modeled by probability theory, the mathematically optimal strategy for revising assumptions after processing new evidence can be described as Bayesian reasoning. Within this approach, the plausibility of competing assumptions is described by prior probabilities of hypotheses $P(H_i)$, the conditional probabilities of evidence by likelihoods of evidence $P(E|H_i)$, and the revised assumptions by posterior probabilities of hypotheses $P(H_i|E)$, according to the Bayes rule:

$$P(H_i|E) \propto P(H_i) \cdot P(E|H_i) \quad \text{(Bayesian update, BUS)} \quad (1)$$

Bayesian update (cf. fig. 1) is a general normative model of decision-making (Mandel, 2014). It has often been applied to judgments in medical situations (base rates of illnesses, sensitivity and specificity of tests, e.g., Gigerenzer & Hoffrage, 1995), and even to model teacher judgments in educational situations (Loibl & Leuders, 2020).

Research has demonstrated that humans' capacity to process information on probabilities in Bayesian reasoning is limited which results in sub-optimal heuristics, such as base-rate neglect (e.g., Kahneman & Tversky, 1996). While often such biased strategies are interpreted as a limitation of human thinking in probabilities (Kahneman & Tversky, 1996), one could also ask whether mathematically suboptimal heuristics may be regarded as adequate and effective reasoning strategies in certain situations (ecological rationality: Simon, 1955; Gigerenzer & Hoffrage, 1995). For instance, Sundh (2019) showed that for calculations with joint probabilities an averaging heuristic was adequate in certain constellations. Similar questions of ecological rationality of heuristics have been investigated with respect to multiple-cue situations using strategies like take-the-best or fast-and-frugal trees (Gigerenzer & Goldstein, 1996; Martignon, Vitouch, Takezawa, & Forster, 2003).

In our study, we focus on a situation that has not been studied in the light of ecological rationality before: Single-cue judgments on multiple (2 or 3) hypotheses with complete but

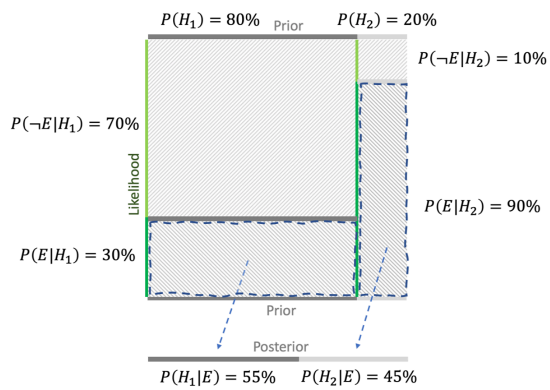

Figure 1: The diagram illustrates the updating of the probability of two mutually exclusive hypotheses H_i. Given the base rates of the hypotheses $P(H_i)$ and the likelihoods of the evidence $P(E|H_i)$, the prior probability $P(H_2) = 20\%$ increases to the posterior probability $P(H_2|E) \approx 45\%$ after evidence E is observed, according to the Bayes rule:

$$P(H_2|E) = \frac{90\% \cdot 20\%}{30\% \cdot 80\% + 90\% \cdot 20\%} \approx \frac{20\%}{45\%} \approx 45\%$$

(ratio of right dashed area to total dashed area) noisy probability information (priors and likelihoods). The noisy estimates of the probabilities (cf. Sundh, 2019) make exact calculations unfeasible. From a cognitive perspective, it seems plausible that humans' mental models of their intuitive estimates on probabilities in real world decisions are analog non-numerical representations (Khemlani, Lotstein, & Johnson-Laird, 2015). Khemlani et al. proposed a computational model assuming primitive analog representations for noisy probabilities and implemented intuitive strategies on processing these non-numerical probabilities in the model. Juslin, Nilsson, & Winman (2009) modeled complex types of reasoning with noisy probabilities (including Bayesian updating, see below). Both computational models were validated with human data.

Against this background, we propose a cognitively plausible heuristic strategy (simpler than Juslin et al., 2009), which amounts to approximately averaging the probability information on prior hypotheses and evidence (APES). We explore the accuracy and ecological rationality of APES for Bayesian reasoning in analog non-numerical settings. In two computational studies, we analyze the relative accuracy of decisions based on APES in a two- and a three-hypotheses situation. Finally, we discuss the scope and ecological rationality of APES by framing it in the context of teacher judgements.

State of Research on Heuristics in Bayesian Reasoning

As outlined above, Bayesian reasoning (for single cues) requires the combination of multiple probabilities in a multiplicative way (e.g., 90% · 20%). However, multiplying probabilities is not intuitive (e.g., 90% of 20%) and therefore it is cognitively demanding (Sundh, 2019). Unsurprisingly, research shows that humans often fail to apply the Bayes rule correctly, even when strongly supported (Gigerenzer & Hoffrage, 1995; Weber, Binder, & Kraus, 2018). More specifically, research has identified often-applied heuristics that lead to biased decisions (e.g., base-rate neglect, Kahneman & Tversky, 1996).

In a systematic analysis on the types of update strategies in the context of numerical Bayes reasoning tasks, Cohen and Staub (2015) showed that most participants' strategies amount to not making use of all sources of information: Most participants estimated the posterior probability based on only one of the multiple provided probabilities or by computing a weighted sum of several, but not all probabilities. In their study, most participants only processed the evidence (*evidence-only strategy, EOS*, cf. Zhu & Gigerenzer, 2006; with variations called representative thinking: Zhu & Gigerenzer, 2006; Fisherian: Gigerenzer & Hoffrage, 1995; inverse fallacy, Villejoubert & Mandel, 2002; likelihood subtraction: Gigerenzer & Hoffrage, 1995). Other participants only took the prior probabilities (*priors only, POS*, cf. base rate only: Gigerenzer & Hoffrage, 1995; also called conservatism: Edwards, 1968; Zhu & Gigerenzer, 2006).

$$P(H_i|E) \propto P(E|H_i) \quad \text{(Evidence only, EOS)} \quad (2)$$
$$P(H_i|E) \propto P(H_i) \quad \text{(Prior only, POS)} \quad (3)$$

The reasoning strategies related to (2) and (3) are characterized by the disregard of information and therefore they are cognitively simpler to perform than BUS (1). Strategies, which combine information additively are discussed in several contexts, such as joint probabilities (Sundh, 2019), conjunctive probabilities (Juslin, Lindskog, & Mayerhofer, 2015), and Bayesian reasoning (Cohen & Staub, 2015; Juslin et al., 2009; Lopes, 1985; Shanteau, 1975). Additive strategies combine all probability information and can approximate multiplicative strategies in some situations. These strategies assume that the individual determine posteriors by a complex weighted sum of all probabilities (e.g., Cohen & Staub, 2015; Juslin et al., 2009):

$$P(H_i|E) = \alpha \cdot P(H_i) + \beta \cdot P(E|H_i) + \gamma \cdot P(E|\neg H_i) \quad (4)$$

It appears rather implausible that such a complex set of information on probabilities and regression weights can be processed intuitively. Therefore, we assume a more cognitively plausible additive strategy (see also Shanteau, 1975):

$$P(H_i|E) \propto \tfrac{1}{2}\big(P(H_i) + P(E|H_i)\big) \quad (5)$$
(Averaging-Prior-Evidence, APES)

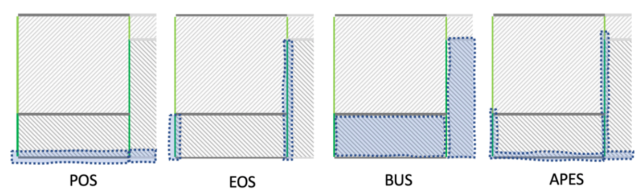

Figure 2: The highlighted areas show how the strategies POS, EOS, BUS, and APES take into account the probability information – priors (grey horizontal bars) and likelihoods (green vertical bars).

Mathematically, APES can be regarded as an approximation, since it qualitatively reflects the "magnitude" of the product in (1), illustrated by the average side length instead of the area of the rectangle in Fig. 2. While POS remains on the prior probabilities (grey horizontal bars) and EOS only uses the likelihoods (green vertical bars), BUS correctly regards the interaction of prior probabilities and likelihoods (multiplication). APES also considers both, but approximates the interaction by averaging the probabilities.

Cognitively, it is simpler to derive an additive average of two magnitudes (e.g., ½·(90% + 20%)) than a multiplicative interaction. This is expressed in the mental model for averaging subjective probabilities (Khemlani et al., 2015) and amounts to a "take-the-middle heuristic".

Empirically, one can find indicators for such additive strategies in the literature on Bayesian reasoning: For instance, the responses of the participants in the study by Cohen and Staub (2015) could be better modeled as an additive combination of multiple probabilities than as a multiplicative combination of probabilities. Shanteau (1975) showed that when updating probability estimations based on (non-informative) evidence, the probability updates of the participants suggest averaging (APES) instead of multiplying (BUS).

Research question

Building on cognitive plausibility and empirical evidence, we argue that in situations without numerical representations one may assume a strategy, which averages probabilities of priors and evidence (APES). This strategy approximates exact Bayesian reasoning by reducing complexity without neglecting information. In two computational explorations, we study the potential effectiveness of APES by investigating the following research question:

For which types of situations (i.e., constellations of prior and likelihood values) can the averaging-priors-and-evidence strategy (APES) be regarded as an adequate and effective approximation of the Bayesian update strategy (BUS) and as a substantial improvement with respect to the information-neglecting strategies prior only and evidence only (POS, EOS)?

Since we embrace a perspective of ecological rationality, we finally discuss the adequacy of the approximative strategy against the background of teachers' diagnostic judgments and thus focus on situations of judging a certain piece of evidence (a student's solution), which is indicative for one of several possible hypotheses (students' misconceptions) (for an empirical investigation of this situation, cf. Loibl & Leuders, 2020).

Methods and Results

Study 1

In the scenario for study 1, we assume a situation with two hypotheses H_1, H_2 and evidence that is indicative for one misconception (high likelihood for H_2) and reduces plausibility for the other misconception (low likelihood for H_1). Accordingly, we explore constellations with the following set of values:

$$P(H_1), P(H_2) \in [0;1], \sum P(H_i) = 1$$
$$P(E|H_1) \in [0.10; 0.40], P(E|H_2) \in [0.60; 0.90]$$

We account for the noisiness of the probability estimates by interpreting the numerical values as centers of approximative intervals, e.g. 25% represents ca. 20-30%.

For the computational simulation of the various strategies, we assume the following cognitive process: The probability information (priors $P(H_i)$ and likelihoods $P(E|H_i)$), and the evidence E are available. The goal is to decide which of the two posterior hypotheses $P(H_i|E)$ has the higher probability. To that purpose one of the following strategies is activated (the ≷-sign meaning "compare and decide for the larger"):

$$P(H_1) \gtrless P(H_2) \quad \text{(POS)}$$
$$P(E|H_1) \gtrless P(E|H_2) \quad \text{(EOS)}$$
$$P(H_1)P(E|H_1) \gtrless P(H_2)P(E|H_2) \quad \text{(BUS)}$$
$$\tfrac{1}{2}[P(H_1) + P(E|H_1)] \gtrless \tfrac{1}{2}[P(H_2) + P(E|H_2)] \quad \text{(APES)}$$

A normalization factor (e.g., $\Sigma_i P(H_i)P(E|H_i)$ for BUS), which is necessary to attain a value of the posterior probability, is not required for a decision between the hypotheses, since it is identical for both sides of the comparison in each strategy.

The simulation and the graphical representation of the results was implemented in Cinderella (Richter-Gebert & Kortenkamp, 2011), a programming environment for numerical calculation and visualization (source code available from the authors).

Results of Study 1

To evaluate and compare the various investigated reasoning strategies, we graphically display the outcome (the decision for a hypothesis) throughout the whole parameter space in a way that makes the phenomena most salient (see Fig. 3-5): Each dot represents a decision based on the respective strategy for a certain set of parameters in the probability space. The decision depends on the comparison of the posterior probabilities for H_1 and H_2. The values of the prior probabilities are represented by the *x*-coordinate. The accuracy of the decision is the corresponding posterior and is represented by the *y*-coordinate. For each value on the *x*-coordinate there is an interval of values on the *y*-coordinate due to the variation of the likelihood values. Numbers ①-③ indicate regions of decisions as explained in the text.

Fig. 3 presents the results of the optimal Bayesian decision (BUS) when evidence is indicative for H_2. In region ①, BUS leads to a decision for H_1 because the prior probability for H_1 is high. When the prior probability for H_2 is middle to high as in region ②, BUS results in a decision for H_2 due to the evidence for H_2. For a certain prior probability constellation (region ③) BUS leads to either H_1 or H_2 depending on the likelihood values.

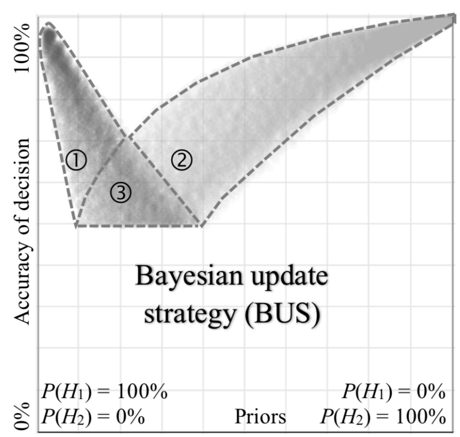

Figure 3: Accuracy of the decision based on the BUS strategy.

BUS is the mathematical optimal update strategy and can be used for decisions by selecting the hypothesis with the highest posterior probability (in case of two hypotheses, the hypothesis with a probability ≥50%). The decision accuracy of the other strategies (EOS, POS, APES) depends on the accuracy of BUS and can never exceed the value for BUS. If a decision coincides with BUS, the accuracy of the decision is the same as for BUS. If the decision deviates from BUS, the accuracy of the decision corresponds to the lower decision accuracy for the opposite hypothesis following BUS (not displayed in Fig. 3). For example, the strategy EOS results in a decision for H_2. For certain prior and likelihood values (④ in Fig. 4), this decision coincides with the optimal decision according to BUS with the accuracy $P_{\text{BUS}}(H_2|E) \geq 50\%$. For other prior and likelihood values, H_1 has the higher probability according to BUS. Here, EOS deviates from BUS (⑤ in Fig. 4) and, thus, the decision based on EOS has the (lower) accuracy $P_{\text{BUS}}(H_2|E) \leq 50\%$.

Fig. 4 presents the comparison of the decisions based on the information neglecting strategies (POS, EOS) and the optimal Bayesian decision (BUS) when evidence is indicative for H_2. The decision of the information neglecting strategies coincides with Bayesian reasoning for certain constellations of priors and likelihoods in regions ④. The decision deviates in the broad regions ⑤ of prior values due to the disregard of evidence (POS) or priors (EOS), resulting in low decision accuracy (i.e., posterior probability according to BUS below 50%) in these regions.

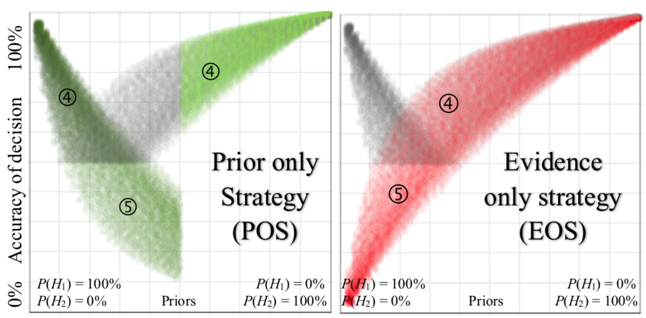

Figure 4: Accuracy of the decision based on information neglecting strategies POS (green, l.h.s.) and EOS (red, r.h.s.) in comparison to the accuracy of BUS (grey in the background).

Fig. 5 presents the comparison of the decision based on the averaging-priors-and-evidence strategy (APES) and the optimal Bayesian decision (BUS) when evidence is indicative for H_2. The decision based on APES coincides with Bayesian reasoning for most constellations of priors and likelihoods as shown by the broad regions ⑥. In the prior region where the BUS decision for H_1 or H_2 depends on the likelihood constellations (compare region ③ in Fig. 3), APES coincides with BUS in region ⑦ and deviates from BUS in region ⑧, depending on the likelihood constellations. Moreover, the deviations only result in small reductions of the accuracy in comparison to the accuracy of the Bayesian decision. The small magnitude of the accuracy reduction is due to the fact that the region with deviations falls in the region with the smallest accuracy of the Bayesian decision, close to 50%.

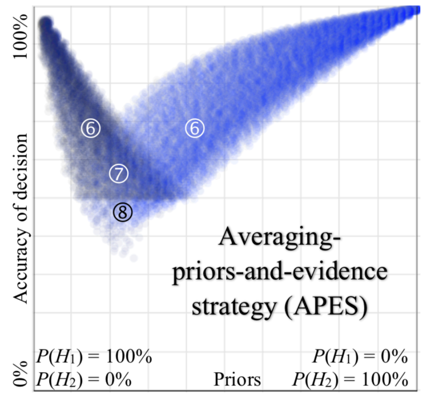

Figure 5: Accuracy of the decision based on the averaging-priors-and-evidence strategy (APES, blue) in comparison to the accuracy of the decision based on BUS (grey).

Study 2

As an extension to the scenario of study 1, we explored the more complex situation of three hypotheses and a situation where the evidence is sensitive and specific for only one of the hypotheses (i.e., high likelihood only for one of the hypotheses, cf. Fig. 6, l.h.s.). We therefore explore situations with the following values:

$$P(H_1), P(H_2), P(H_3) \in [0;1], \sum P(H_i) = 1$$
$$P(E|H_1), P(E|H_2) \in [0.10; 0.40], P(E|H_3) \in [0.60; 0.90]$$

The simulation algorithm of the strategies corresponds to study 1: A decision for one of the three hypotheses is reached by deciding for the highest posterior probability according to each strategy. The representation of the results, however, has to be adapted to the affordances of the higher dimensionality of three hypotheses. To that purpose, we use barycentric homogeneous coordinates. The diagram, which we call "hypothegon" (cf. Fig. 6, r.h.s.; de Finetti, 2017; Leuders & Loibl, 2020; Jøsang, 2016), extends the hypothesis line [0,1] for two hypotheses to a triangular prior space for three hypotheses.

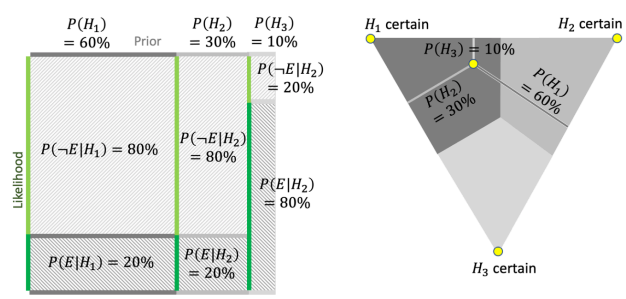

Figure 6: A situation with three hypotheses and their likelihoods (low likelihood 20% for H_1 and H_2, high likelihood 80% for H_3) with respect to evidence E (l.h.s). Any set of prior probabilities (e.g., 60%, 30%, 10%) can be regarded as convex coordinates for a unique locus within a triangle ("hypothegon") (r.h.s.).

Results of Study 2

As outcome of the various strategies, we again display the accuracy of each decision, depending on the value of the prior probabilities. In order to attain a more pronounced geometry, we chose to only display the results for a fixed likelihood (80%), keeping in mind that the likelihood interval leads to an interval of the accuracy value for fixed prior probabilities.

Fig. 7 gives a comprehensive picture of the averaging-prior-and-evidence strategy (APES, blue) compared to the Bayesian strategy (BUS, grey) when evidence is indicative for H_3. In most regions, both strategies *coincide* with high decision accuracy in regions ① with a large prior probability for one of the hypotheses (H_1, H_2, or H_3) and medium decision accuracy for region ② with similar prior probabilities for all hypotheses ($H_1 \approx H_2 \approx H_3 \approx 30\%$). The only prior region in which APES *deviates* from BUS, is region ③ with low prior probability for H_3 and similar prior probabilities for H_1 and H_2 ($H_1 \approx H_2 \approx 50\%$). In our setting, the evidence is contra-indicative for H_1 and H_2 with ambiguous likelihoods for these hypotheses. This ambiguity comes into effect only in regions with similar and rather high prior probabilities. Due to the ambiguity the accuracy of APES drops from 70%-50% to 50%-0% in region ③. Further variation of the likelihood values (not displayed here) does not alter the general picture but only slightly increases the region ③ in which APES deviates from BUS.

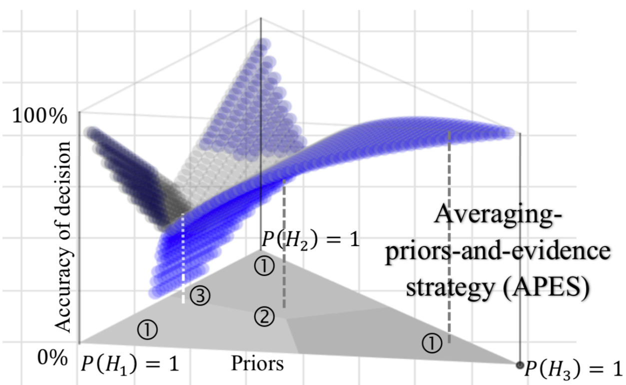

Figure 7: Accuracy of the decision based on the averaging-priors-and-evidence strategy (APES, blue) in comparison to the accuracy of BUS (grey) with regard to three hypotheses. The prior probabilities are coordinates within the triangle (cf. Fig. 6).

In order to compare the deviations of the decision based on the heuristic strategies EOS, POS, and APES from the ideal Bayesian decision (BUS), we calculate the differences in decision accuracy (in %) and display the resulting "error distribution" over the whole prior space (hypothegon), for fixed likelihoods (20%, 20%, 80%) (Fig. 8). For the information neglecting strategies EOS and POS, there are – similar to study 1 with two hypotheses (Fig. 4) – large deviations for EOS (displayed in red) in region ① with extreme priors contrary to the evidence (EOS decision H_3, BUS decision H_1 or H_2) and large deviations for POS (displayed in green) in region ② with priors slightly contrary to the evidence (POS decision H_1 or H_2, BUS decision H_3).

In contrast, for APES (displayed in blue) the deviations are much smaller and restricted to region ③ with extreme priors contrary to the evidence and ambiguous likelihoods for H_1 and H_2. For more extreme likelihood parameters (e.g., 10%, 10%, 90%) this region is closer to the extreme boundary and for less extreme likelihood parameters (e.g., 40%, 40%, 60%) it approaches the middle but with only very small values of deviation (<10%, not displayed here). As shown in Fig. 8, there are no deviations for any strategy, when priors and evidence suggest identical decisions.

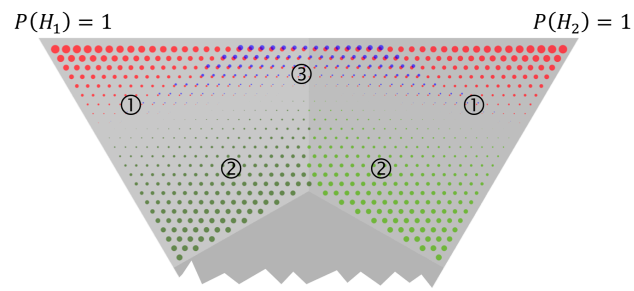

Figure 8: Deviations of decision accuracy for EOS, POS, and APES from the ideal Bayesian decision. The size of the dots corresponds to the magnitude of the deviations with the largest dots (e.g., in the corners) corresponding to a 100% deviation.

Discussion

In our study, we investigated heuristic strategies in Bayesian reasoning with respect to their accuracy, depending on the set of values of prior probabilities of the hypotheses and likelihoods of the evidence. In addition to the mathematically optimal Bayesian strategy (BUS) and the well-documented heuristic strategies, that neglect either evidence or prior information (EOS, POS), we proposed an averaging strategy (APES) and underpinned its plausibility based on cognitive and empirical arguments from literature (e.g., Cohen & Staub, 2015; Khemlani et al., 2015; Shanteau, 1975).

The exploration of a broad parameter space (two and three hypotheses with priors from [0;1], evidence with likelihoods from [0.6;0.9]) yielded the following insights:

- Processing of evidence only (EOS, prior neglect) or priors only (POS) results in low accuracy compared to the Bayesian update strategy (BUS) in broad regions of prior probabilities. These deviations occur when priors and evidence suggest divergent decisions (⑤ in Fig. 4; ① & ② in Fig. 8).
- Averaging prior and evidence probabilities (APES) is a good approximation for Bayesian reasoning, leading to identical decisions for most values of priors and likelihood (⑥, ⑦ in Fig. 5; ① & ② in Fig. 7 & Fig. 8).

- Deviations of APES from BUS remain small and occur in confined regions of prior values. We found two situations (ranges of values) where the decision based on APES deviates from the BUS decision: (1) similar values of competing posteriors (⑧ in Fig. 5), which is also the most inconclusive (i.e., least accurate) situation in exact Bayesian reasoning, (2) ambiguous hypotheses (here: hypotheses with similar low likelihoods) sharing high prior probabilities (③ in Fig. 7 & 8).

From a perspective of ecological rationality, heuristic strategies possibly constitute viable and efficient cognitive strategies in certain situations (cf. Simon, 1955; Gigerenzer & Hoffrage, 1995). To investigate the heuristic strategies with regard to their ecological rationality, we specify the situational framing as teachers' diagnostic judgments on students' potential misconceptions (with certain prior probabilities) based on responses to tasks (evidence). For instance, what is the plausibility of the hypothesis H_2 that a student has misconception 2 (e.g., "more decimals make smaller numbers") after a response to a task (e.g., "1.52 < 1.4")? A teacher ideally considers all information (as subjective probabilities not necessarily represented numerically) and processes this information to arrive at a decision on the misconception with the highest posterior probability. However, given the complexity of the situation, it is likely that a teacher applies one of the heuristic strategies. For the case of teachers' decisions, we put forward the following interpretations:

- As teachers use diagnostic tasks in order to receive evidence about their students' skills and misconceptions, teachers are not likely to ignore this evidence. We therefore argue that teachers do not apply a priors-only strategy (POS) when evidence is available.
- When teachers apply an evidence-only strategy (EOS), they focus on students' errors and tend to react immediately with instruction (Herppich, Wittwer, Nückles, & Renkl, 2016; Phelps-Gregory & Spitzer, 2018). This may be an inaccurate diagnostic decision for misconceptions with low base rate (and inefficient considering restricted instructional time). However, when applied as screening for further diagnostic interaction, the strategy may be appropriate.
- While Bayesian reasoning is computationally rather complex, a simple averaging strategy (APES) seems not only cognitively plausible, but also appropriate and feasible in situations with quick on-the-fly assessment. Our results indicate that they lead to optimal (Bayesian) decisions in most cases, and only deviate in situations of ambiguity (similar posteriors). In these cases, an appropriate decision for teachers would be to *not* decide on one misconception or the other but to resume assessment.

This analysis gives strong support to consider averaging and considering all information (APES) as a promising heuristic strategy for Bayesian decision situations in general, and ecologically valid for teachers' diagnostic judgments. However, in spite of ecological rationality, we do not posit that this plausibility implies actual prevalence: We have no information on teachers' application of such a strategy during teaching. Furthermore, teachers' diagnostic cognitive processes are far more complex than the focus of our analysis (e.g., single- vs. multiple-cues judgments) and require more comprehensive models (Herppich et al., 2018; Loibl, Leuders, & Dörfler, 2020).

Of course, it would be interesting to further investigate the heuristic strategies within more complex situations (multiple cues, further sets of values for sensitivity of evidence), and to better understand the role of ambiguity. Also, the computational model here does not explicitly reflect the noisiness of analog non-numerical mental representations of subjective probabilities (cf. research on magnitude and analogue representations: Gallistel, 2011; Khemlani et al., 2015; Leibovich, Katzin, Harel, & Henik, 2017 or models for uncertainty in probability estimation: Jøsang, 2016).

However, we consider the most pressing aim for further research to empirically ascertain the validity of the averaging strategy (APES) in non-numerical situations, since empirical support is still rare. Deriving empirical evidence for APES in non-numerical situations may be challenging due to the small deviations of APES from BUS. A promising route to this goal is to extend the model from deciding on the largest posterior to estimating posterior values. Following this route, we gained first experiences with intervention studies in which we systematically prompted the processing of all probability information similar to APES in the context of teacher judgments (Loibl & Leuders, 2020).

References

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

De Finetti, B. (2017). *Theory of Probability: A Critical Introductory Treatment*. New York, NY: John Wiley & Sons.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal Representation of Human Judgment* (pp. 17-52). New York, NY: Wiley.

Gallistel, C.R. (2011). Mental magnitudes. In S. Dehaene, E.M. Brannon (Eds.), *Space, time and number in the brain: Searching for the foundations of mathematical thought* (pp. 3-12). London: Elsevier.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review, 103*(4), 650.

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

Herppich, S., Wittwer, J., Nückles, M., & Renkl, A. (2016). Expertise amiss: Interactivity fosters learning but expert tutors are less interactive than novice tutors. *Instructional Science, 44,* 205-219.

Herppich, S., Praetorius, K., Förster, N., Glogger-Frey, I., Karst, K., Leutner, D., Behrmann, L., Böhmer, M., Ufer, S., Klug, J., Hetmanek, A., Ohle, A., Böhmer, I., Karing, C., Kaiser, J., & Südkamp, A. (2018). Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teaching and Teacher Education, 76*, 181-193.

Jøsang, A. (2016). Generalising Bayes' theorem in subjective logic. *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)* (pp.462-469). Baden-Baden.

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

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

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

Khemlani, S. S., Lotstein, M., & Johnson-Laird, P. N. (2015). Naive probability: Model-based estimates of unique events. *Cognitive Science, 39*(6), 1216-1258.

Leibovich, T., Katzin, N., Harel, M., & Henik, A. (2017). From 'sense of number' to 'sense of magnitude' – the role of continuous magnitudes in numerical cognition. *Behavioral and Brain Sciences, 40*, E164.

Leuders, T., & Loibl, K. (2020). Processing probability information in non-numerical settings – teachers' bayesian and non-bayesian strategies during diagnostic judgment. *Frontiers in Psychology,* doi: 10.3389/fpsyg.2020.00678

Loibl, K., Leuders, T., & Dörfler, T. (2020). A Framework for Explaining Teachers' Diagnostic Judgements by Cognitive Modeling (DiaCoM). *Teaching and Teacher Education*, 91. doi: 10.1016/j.tate.2020.103059

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

Mandel, D. R. (2014). The psychology of bayesian reasoning. *Frontiers in Psychology*, 5, 1144.

Martignon, L., Vitouch, O., Takezawa, M., & Forster, M. R. (2003). Naive and yet enlightened: From natural frequencies to fast and frugal decision trees. In D. Hardman & L. Macchi (Eds.), *Thinking: Psychological perspective on reasoning, judgment, and decision making* (pp. 189-211). Chichester, UK: Wiley.

Phelps-Gregory, C. M., & Spitzer, S. M. (2018). Developing prospective teachers' ability by classroom intervention: Replicating a classroom intervention. In T. Leuders, J. Leuders, & K. Philipp (Eds.), *Diagnostic Competence of Mathematics Teachers: Unpacking a complex construct in teacher education and teacher practice* (pp. 223-240). New York: Springer.

Richter-Gebert, J., & Kortenkamp, U. H. (2000). *User manual for the interactive geometry software cinderella*. Springer Science & Business Media.

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

Simon, H. A. (1995). A behavioral model of rational choice. *The Quarterly Journal of Economics, 69*(1), 99-118.

Sundh, J. (2019). The Cognitive Basis of Joint Probability Judgments. Processes, Ecology, and Adaption. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences, 166*. Uppsala: Acta Universitatis Upsaliensis.

Villejoubert, G., & Mandel, D. R. (2002). The inverse fallacy: an account of deviations from Bayes' theorem and the additivity principle. *Memory & Cognition, 30*(2), 171-178.

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

Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: The role of representation in mental computation. *Cognition, 98,* 287-308.