What's in a sample? Epistemic uncertainty and metacognitive awareness in risk taking

Sebastian Olschewski [a,b,*], Benjamin Scheibehenne [c]

[a] Department of Psychology, University of Basel, Switzerland
[b] Warwick Business School, University of Warwick, United Kingdom
[c] Department of Economics and Management, Karlsruhe Institute of Technology, Germany

A B S T R A C T

In a fundamentally uncertain world, sound information processing is a prerequisite for effective behavior. Given that information processing is subject to inevitable cognitive imprecision, decision makers should adapt to this imprecision and to the resulting epistemic uncertainty when taking risks. We tested this metacognitive ability in two experiments in which participants estimated the expected value of different number distributions from sequential samples and then bet on their own estimation accuracy. Results show that estimates were imprecise, and this imprecision increased with higher distributional standard deviations. Importantly, participants adapted their risk-taking behavior to this imprecision and hence deviated from the predictions of Bayesian models of uncertainty that assume perfect integration of information. To explain these results, we developed a computational model that combines Bayesian updating with a metacognitive awareness of cognitive imprecision in the integration of information. Modeling results were robust to the inclusion of an empirical measure of participants' perceived variability. In sum, we show that cognitive imprecision is crucial to understanding risk taking in decisions from experience. The results further demonstrate the importance of metacognitive awareness as a cognitive building block for adaptive behavior under (partial) uncertainty.

1. Introduction

In a fundamentally uncertain world, learning by sampling information from the environment and acting on experiences are crucial building blocks of adaptive behavior (Dall, Giraldeau, Olsson, McNamara, & Stephens, 2005; Fiedler, 2000; March, 2010). To understand and predict such behavior, it is important to characterize both the information provided by the environment and the ability of an organism to absorb and process this information (Lieder & Griffiths, 2020; Simon, 1955, 1959, 1986). Because physical and cognitive abilities of organisms are bounded, owing to inevitable biological constraints, adaptive behavior requires taking one's own limitations into account. Being adaptive in that sense ensures that one will not waste resources on unlikely or impossible goals but will instead pursue realistic and rewarding objectives.

Uncertainty can be aleatory or epistemic (Fox & Ülkümen, 2011; Soll, Palley, Klayman, & Moore, in press; Tannenbaum, Fox, & Ülkümen, 2017; Walters, Ülkümen, Tannenbaum, Erner, & Fox, 2023). Aleatory uncertainty, often quantified as the variance or standard deviation (SD), is an unchangeable property of the environment (Hacking, 2006). In contrast, epistemic uncertainty emerges

---

* Corresponding author at: Missionsstrasse 62A, Center for Economic Psychology, University of Basel, Switzerland.
   E-mail address: sebastian.olschewski@unibas.ch (S. Olschewski).

from the organism's available information and processing capacity (Enke & Graeber, 2023; Lee & Usher, 2023; Heath & Tversky, 1991). This epistemic uncertainty can further be divided into so-called Brunswikian and Thurstonian uncertainty (Juslin & Olsson, 1997). Brunswikian uncertainty refers to a lack of information or a less than perfect correlation between the information and the to-be-predicted states of the world. Hence, it refers to uncertainty that results from sampling noise, which is different from aleatory uncertainty that refers to distribution/ population parameters. For example, there is Brunswikian uncertainty around the estimate of the expected value (EV) from a sample mean. This uncertainty can be reduced through sampling more. In contrast, Thurstonian uncertainty refers to cognitive imprecision in the representation and integration of information. This uncertainty originates in the human cognitive architecture and can increase uncertainty in decision making beyond the other statistical measures of uncertainty.

1.1. Brunswikian uncertainty and Bayesian updating

When making decisions based on information sampled from the environment, the risk that an organism faces consists of both aleatory and epistemic uncertainty. Most research on risk taking when sampling information focuses on aleatory uncertainty as the principal source of risk (e.g., Barron & Erev, 2003), yet a comprehensive theory of adaptive risk taking under (partial) uncertainty must also consider epistemic uncertainty, which is the focus of our studies.

We start by first examining Brunswikian uncertainty. In cases where one needs to infer population characteristics from (relatively small) samples, Brunswikian uncertainty refers to the fact that the statistical properties of the observed sample (e.g., its mean) can differ from the population characteristics. This statistical uncertainty can be reduced through information sampling. For example, an agent observing only two reward samples faces more statistical uncertainty about the expected reward than an agent observing 20 samples from the same population. This statistical uncertainty can be quantified through Bayes's theorem, which is often portrayed as the rational benchmark in inference and preference tasks (e.g., Kuhnen, 2015; Navajas et al., 2017; Vul, Goodman, Griffiths, & Tenenbaum, 2014; see also Tauber, Navarro, Perfors, & Steyvers, 2017). Recent research on Bayesian cognition quantified the statistical uncertainty as the Bayesian posterior variability of the statistic of interest (Xiang, Graeber, Enke, & Gershman, 2021). With respect to the expected reward of a payoff distribution, the posterior variability of the estimate increases with the variability of information (i.e., its SD) and decreases with the square root of the amount of information (i.e., the sample size).

Yet, Bayesian posterior variability is only helpful for understanding risk taking to the extent that all information is integrated precisely. This is not a constraint for Bayesian updating in a mathematical framework, but humans also face Thurstonian uncertainty; their precision when integrating information is limited by fundamental cognitive constraints (Cowan, 2001; Hahn, 2014; Khaw, Li, & Woodford, 2021; Miller, 1956; Simon, 1986; Woodford, 2020). More generally, the question of whether humans can perform Bayesian calculus is an important topic for the debate on Bayesian updating as a descriptive model of human cognition (Benjamin, 2019;
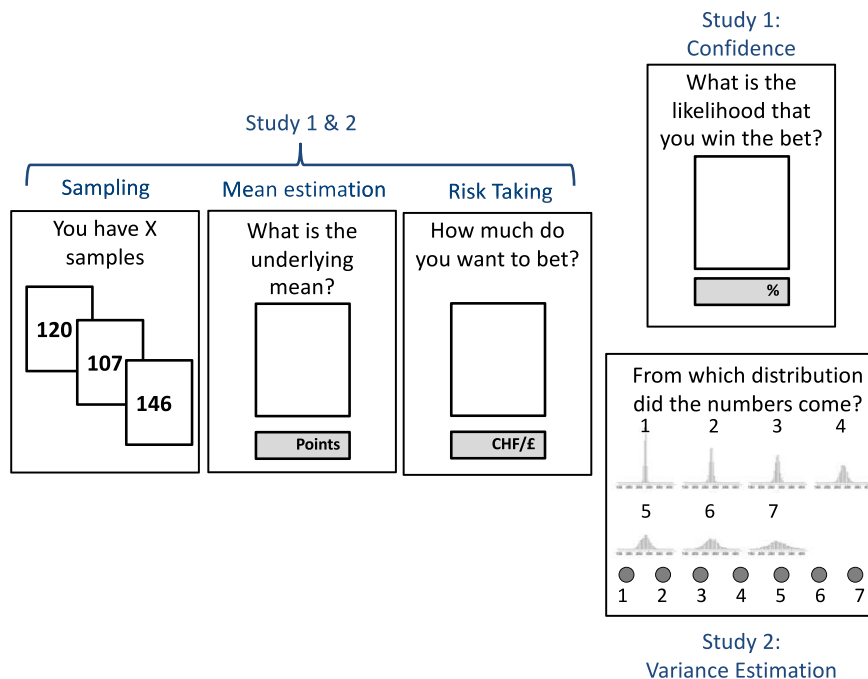

Fig. 1. *Stylized screenshots of one trial of the experimental design.* First, participants sampled exactly *x* times from a distribution, and *x* was either 4, 6, 16, or 24. Second, participants had to estimate the mean of the underlying distribution on the basis of the samples they saw to measure the accuracy of information integration. Third, participants decided how much to bet on the accuracy of their mean estimate. We used their answers as a measure of risk taking. Fourth, participants either gave their estimated probability of winning the bet as a measure of confidence (Study 1) or decided from which of seven histograms with varying variance the samples most likely were drawn as a measure of perceived variance (Study 2).

Charness & Levin, 2005; Dasgupta, Schulz, Tenenbaum, & Gershman, 2020; Holt & Smith, 2009; Oaksford & Chater, 1994; Tversky & Kahneman, 1971). To the degree that information is processed imprecisely, the epistemic uncertainty a human agent faces will be affected by both Brunswikian uncertainty, as represented by the Bayesian posterior variability, and Thurstonian uncertainty due to cognitive imprecision. Therefore, adaptive risk taking under epistemic uncertainty should be guided not just by Bayesian posterior variability alone, but also by the degree of imprecision in information processing. This assumes, however, that decision makers have metacognitive insight into their cognitive limits when integrating information in a given environment (Enke & Graeber, 2023). In other words, they need to know that they do not know.

1.2. Thurstonian uncertainty and decisions from experience

To empirically test if decision makers adapt to the additional Thurstonian uncertainty that arises from imprecise information processing, we developed an experimental task that we applied across two preregistered studies with a student and a general population online sample. Our task is based on the decisions from experience (DfE) experimental paradigm, where participants sequentially sample outcomes from otherwise unknown choice alternatives before making one consequential choice (Hertwig & Erev, 2009). In DfE tasks, the differences in risk taking compared to decisions from description, where all relevant information about probabilities and outcomes is summarized, highlight the importance of information presentation and sampling for subsequent behavior (Barron & Erev, 2003; Hertwig, Barron, Weber, & Erev, 2004; Ludvig & Spetch, 2011; Spitzer, Waschke, & Summerfield, 2017; Tsetsos et al., 2012, 2016; Wulff, Mergenthaler-Canseco, & Hertwig, 2018). In typical DfE tasks, risk taking is subject to both aleatory and epistemic uncertainty (the Brunswikian and Thurstonian uncertainty in estimating the outcome distributions). Our paradigm is designed to experimentally disentangle these sources of uncertainty. In particular, we isolate the epistemic part of the uncertainty by asking decision makers to place a monetary bet on their accuracy when estimating the EV of a distribution. Brunswikian uncertainty can be quantified theoretically as the Bayesian posterior variability that derives from the distribution's SD and the sample size. With these mathematical relations we can hold the posterior variability constant while varying the SD and sample size. In this comparison—holding posterior variability constant while varying the SD and sample size—observed differences in estimation accuracy can be attributed to cognitive imprecision in the representation and integration of information. This is the basic setup for our experimental design that we outline in more detail in Section 3.1 (see Fig. 1 for a schematic of the task).

To the degree that participants in our experiments would face Thurstonian uncertainty due to imprecise information processing and integration, we expected the epistemic uncertainty of their estimates to be higher than the Brunswikian uncertainty predicted by Bayesian models that integrate all information in an optimal way. Moreover, we expected that participants' epistemic uncertainty would depend on one or both of the following two factors: First, we expected that high variability of sampled information would increase Thurstonian uncertainty of the estimate relative to low variability, because for human decision makers, more cognitive imprecision arises when estimating the midpoint of numbers further apart than closer together (Laestadius, 1970; Navajas et al., 2017; Olschewski, Newell, Oberholzer, & Scheibehenne, 2021; Peterson & Beach, 1967; Spencer, 1963; Wolfe, 1975). Second, we expected that Thurstonian uncertainty of participants' estimates would increase with the number of observed samples because of cognitive imprecision when integrating high amounts of information (Beach & Swenson, 1966; Obrecht, Chapman, & Gelman, 2007; Spencer, 1963; but see Brezis, Bronfman, & Usher, 2015). Note that this prediction refers only to the Thurstonian part of uncertainty; the statistical impact of sample size on the Brunswikian uncertainty moves in the opposite direction.

Moreover, cognitive imprecision when integrating information will also affect the accuracy of other judgments about an outcome distribution, such as its perceived variance. Judgments about the variance of an outcome distribution are important for risk taking as they could further affect perceived epistemic uncertainty in a Bayesian framework. These judgments could be distorted by sample size, as the uncorrected sample SD increases with higher sample size, and previous studies showed that human decision makers might base their judgment on the uncorrected sample SD (Kareev, Arnon, & Horwitz-Zeliger, 2002; Konovalova & Le Mens, 2020).

1.3. Metacognitive awareness and confidence

For either of the above-discussed cognitive imprecisions in information processing to guide risk-taking behavior, decision makers require a metacognitive awareness of these limitations. One way to empirically measure metacognitive awareness is through subjective confidence judgments. These confidence judgments have themselves become an important target of psychological theories and have been shown to predict observable behavior in several contexts (e.g., Folke, Jacobsen, Fleming, & De Martino, 2016; Johnson & Fowler, 2011; Lebreton, Abitbol, Daunizeau, & Pessiglione, 2015; Meyniel, Sigman, & Mainen, 2015; Rosenbaum, Glickman, Fleming, & Usher, 2022). For example, humans can qualitatively distinguish between easy and difficult inference problems, and this is also reflected in their confidence (e.g., Boundy-Singer, Ziemba, & Goris, 2023). Yet, prior research also portrayed people as overconfident and quantitatively not well calibrated (Johnson & Fowler, 2011; Malmendier & Tate, 2015; Walters, Fernbach, Fox, & Sloman, 2017; but see Erev, Wallsten, & Budescu, 1994). In particular, people show signs of overprecision; that is, they attribute too much probability mass to the narrow region of their subjective beliefs. For example, their stated confidence intervals, quantiles, or constructed distributions include the true value less often than normatively demanded (Budescu & Du, 2007; Moore & Healy, 2008; Soll et al., in press). Thus, an important part of our research goal was to examine whether people have an adequate representation of epistemic uncertainty when sampling numeric information in a DfE task.

To sum up, to the extent that processing of sampling information is tarnished by cognitive imprecision, adaptive behavior implies that decision makers base their risk taking on their full epistemic uncertainty rather than only on Brunswikian uncertainty modeled by the Bayesian yardstick that assumes perfect information integration.

2. A standard Bayesian model as a quantitative benchmark for Brunswikian uncertainty

We now formalize the Brunswikian uncertainty in a Bayesian updating model for our betting task as a benchmark to compare it against participants' behavior. We assume that a decision maker wants to estimate the EV θ of an underlying distribution based on a sample of observations *y* from that distribution. The decision maker has an uninformative prior about the EV estimate and assumes that the samples come from a normal distribution with a known SD, σ. In this case the posterior distribution of the estimate can be approximately calculated as (Gelman, Carlin, Stern, & Rubin, 1995):

$$p(\widehat{\theta}|y) \approx N\left(\overline{y}, \frac{\sigma^2}{n}\right) \tag{1}$$

Importantly, the SD of the posterior distribution is σ divided by the square root of the sample size *n*. This is the Bayesian posterior variability for the EV estimate under the above-mentioned assumptions. It also corresponds to the standard error of the mean in frequentist statistics.

This analytical solution allows one to construct pairs of combinations that vary in SD and sample size *n* but have the same Bayesian posterior variability as defined in Eq. (1). Table 1 gives an overview of these combinations that provide the basis of our experimental studies. For example, in the upper left cell of Table 1, a sequence with $n = 4$ and $SD = 5$ leads to $SD_{posterior} = 2.50$ and so does the sequence in the cell with $n = 16$ and $SD = 10$. Thus, this model of Bayesian updating predicts equal uncertainty when estimating the EV of these sequences. However, according to our hypotheses from above, the combination with higher SD and higher sample size *n* can lead to different amounts of additional Thurstonian uncertainty for human decision makers. This in turn can affect the risk that decision makers face when betting on the accuracy of their estimate.

3. Study 1

In this study we first tested how the SD and the sample size of a sequence of numbers affect the epistemic uncertainty of human decision makers when estimating the EV of a distribution. Our main goal was then to examine to what extent decision makers take their epistemic uncertainty into account in their subsequent risk taking from experience. Finally, we also examined to what degree reported confidence aligns with the observed epistemic uncertainty.

3.1. Method

3.1.1. Study design

Participants in front of a computer screen repeatedly drew a fixed set of samples from an underlying number distribution. There were 16 trials with different number sequences and the order of the 16 different sequences was randomized for each participant. Sampling took place by pressing the spacebar or Enter key on the keyboard. After each press, participants saw the sampled number for 500 ms. Once the number disappeared, the next sample could be drawn at any time. Following each sampling phase, participants estimated the EV of the respective underlying distribution. We explained the EV as the long-run average when sampling many numbers. Next, participants placed a bet that their estimate was within the range of ± 5 points of the EV. If they were correct their stake was doubled; otherwise they lost their stake. Participants could choose between five stakes from 1 to 5 Swiss francs (CHF) in integers. As a third task we estimated participants' confidence in their estimates. For this, participants stated their perceived probability on a scale of 0 % to 100 % that their previous estimate was within the range of ± 5 points of the EV and that they thus would win their bet. There was no feedback on any of these tasks during the experiment.

3.1.2. Stimuli creation

We implemented a fully balanced experimental design in which we systematically varied the SD of the underlying normal distributions ($SD = 5, 10, 20,$ or 40) and the number of samples ($n = 4, 6, 16,$ or 24) presented to participants in each trial. This yielded 16

Table 1
Standard Deviation and Sample Size Combinations Used as Stimuli in Studies 1 and 2.

Sample size	Standard deviation			
	5	10	20	40
4	2.50ₐ	5.00_c	10.00_e	20.00*
6	2.04_b	4.08_d	8.16_f	16.32*
16	1.25*	**2.50ₐ**	**5.00_c**	**10.00_e**
24	1.02*	**2.04_b**	**4.08_d**	**8.16_f**

Note. The cells show the standard deviation (SD) of the Bayesian posterior of the expected value (EV) estimate with known distributional SD and uninformative prior according to Eq. (1). This posterior variability is a measure of uncertainty and directly translates to the expected probability of winning a bet, defined as an estimate within a range of ±5 points from the EV. We extracted six pairs that had identical Bayesian posterior variability but different combinations of distributional SD and sample size *n*, indicated by identical subscripts. Cells in italic belong to the small *SD & n* group, and cells in bold belong to the high *SD & n* group. Cells with an asterisk (*) had no comparison trial and were included to examine the main effects of SD and sample size.

cells, of which six pairs had identical Bayesian posterior variability according to the analytic solution presented in Eq. (1), but either relatively low or relatively high *SD & n* (Table 1). In total, the low *SD & n* group consisted of trials with $n = 4$ or 6 and *SD* = 5, 10, or 20 and the high *SD & n* group consisted of trials with $n = 16$ or 24 and *SD* = 10, 20, or 40.

For each of the 16 *SD & n* combinations in Table 1 we constructed 60 number sequences with EVs between 280 and 457 in steps of 3. To ensure that the presented sequences were representative of the respective underlying distribution, we forced each sequence to have an SD that deviated maximally 5 % from the target SD. The mean of each sequence was allowed to vary by 0.49 from the EV. That way, in later analyses we control for the statistical effect of sample size on the accuracy of the sample mean with respect to the EV. Finally, we excluded sequences with numbers below 1 and above 999. Each participant saw one sequence randomly drawn from the pool of 60 sequences for each of the combinations.

3.1.3. Procedure and incentives

The experiment was programmed on a computer with PsychoPy (Peirce, 2009) and was conducted in individual sessions in a behavioral laboratory at the University of Geneva. In this and Study 2, all instructions were presented on the computer screen at participants' own pace. Following the instructions, participants saw three questions with three answer options each that probed their understanding of a sequence's mean, the understanding of the difference between a sequence's mean and the distribution's EV, and the understanding of the betting mechanism.

Participants were paid a bonus calculated on the basis of a randomly selected trial for which we compared their estimate against the EV. If their estimate was within a range of ± 5 points of the EV, participants won their bet; otherwise they lost. The final bonus consisted of the amount that had not been bet plus double the stake when the participant won or nothing when the participant lost. There was a fixed participation fee of CHF 20.

3.1.4. Participants and sample size

We calculated our sample size with a medium effect size of $d = 0.5$. A sample size of 54 gave us a 95 % power to detect an effect size of $d = 0.5$ at a significance level of 5 % with a two-sided paired *t* test on the regression coefficient against the null hypothesis of a coefficient of zero. To account for possible exclusions following our preregistered exclusion criteria (see details in Section 3.1.5), we invited 60 participants from the lab participant pool consisting of students from different departments. Of the 56 participants used for analysis, 35 were female, 21 were male, and the mean age was 22 years (range 17–49 years). We collected informed consent from all participants and complied with the lab's ethics regulations.

3.1.5. Preregistration and exclusion

The experimental design and the data analysis were preregistered at OSF: https://osf.io/36rsh/. In accordance with our preregistered exclusion criteria, we removed all observations from trials where estimates were further away than ± 5 SDs of the EV. We removed two participants who made more than two errors in the three comprehension questions, one who gave more than five answers beyond ± 5 SDs of the EV, and one who did both. All data, analysis code, and research materials are available at https://osf.io/36rsh/. All presented regression results were from preregistered mixed-effect regressions with random participant intercepts and slopes but no interactions and were estimated with the lme4 (Bates, Mächler, Bolker, & Walker, 2015) and lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017) packages in R (R Core Team, 2021) unless stated otherwise.
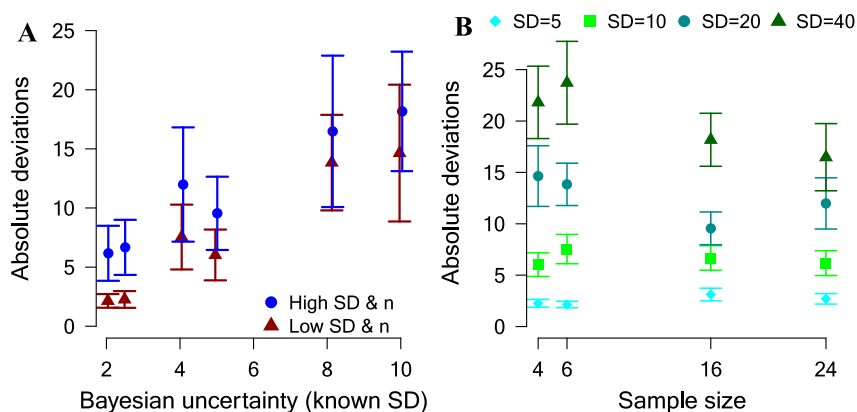

Fig. 2. *Estimation accuracy in Study 1.* Absolute deviations between estimates and expected values compared to predicted Bayesian uncertainty with known standard deviation (SD; Panel A) and compared to sample-generating SD and sample size (Panel B) for Study 1. In Panel A, only trials from the six pairs with identical Bayesian uncertainty with known SD are depicted and high and low refer to the SD and sample size (high: *SD* = 10, 20, or 40 and $n = 16$ or 24; low: *SD* = 5, 10, or 20, and $n = 4$ or 6; see Table 1). Error bars are 95 % confidence intervals.

3.2. Results

3.2.1. Estimation accuracy and cognitive limitations

To empirically measure the epistemic uncertainty that participants faced, we calculated the absolute deviation of their estimates from the EVs. We compared this measure of estimation accuracy for trials that entailed the same Bayesian posterior variability but differed in SD & n (i.e., the cells with identical subscripts in Table 1). This implies that in these trials, Bayesian updating predicted the same estimation accuracy. Results show that participants' deviations from the EVs within these pairs were significantly lower in the low (M = 7.96, 95 % confidence interval [CI] [6.47, 10.51]) compared to the high SD & n condition (M = 11.56, 95 % CI [8.72, 15.94]; regression: b = –3.90, 95 % CI [–6.47, –1.32], p =.004, d = –0.23). This difference indicates that participants' information processing in the high SD & n condition was less accurate and their empirical win rate of the bet was lower (42 %) than in the low SD & n condition (60 %). Fig. 2A shows that this difference in estimation accuracy occurred across all pairs with equal Bayesian posterior variability.

To examine which feature of the number sequences caused the difference in estimation accuracy, we conducted regression analyses across all 16 trials of the fully balanced design (Fig. 2B). The analysis shows that deviations from the EV increased with higher SD (b = 0.47, 95 % CI [0.36, 0.59], p < .001). Given that the number sequences were designed such that the sample mean was close to the distribution's EV across all conditions, this result indicates that participants integrated highly variable information less accurately. In contrast, we did not find evidence for an effect of sample size on participants' deviations from the EV (b = –0.12, 95 % CI [–0.26, 0.01], p = .074). Considering again that the sample mean was close to the distribution's EV for all sample sizes, that is, that we can disregard sampling noise, we found no evidence that participants differed in their accuracy when processing samples for low and high numbers of samples.

Participants' estimates were systematically lower than the EV. A regression on the directed deviations between the estimates and the EVs with just an intercept yielded b = –2.56, 95 % CI [–4.20, –0.92], p = .003, d = –0.12. This systematic underestimation deviated from the Bayesian predictions under uninformative mean priors and decreased overall estimation accuracy above and beyond the effect of SD. Note that this bias, although significant, is relatively small, and the risk of winning one's bet is mainly determined by the estimation accuracy, that is, the expected (absolute) deviation of the estimates from the EVs reported at the beginning of this section.

3.2.2. Adaptive risk taking

Participants' estimation accuracy and thus the epistemic uncertainty they faced in the subsequent betting task differed systematically from the benchmark measure of Bayesian posterior variability. This raises the question of whether participants' betting behavior aligned with this uncertainty. The data indicate that this was indeed the case: Betting stakes were higher in the low SD & n (M = 3.10, 95 % CI [2.83, 3.34]) compared to the high SD & n (M = 2.60, 95 % CI [2.40, 2.82]) condition (regression: b = 0.50, 95 % CI [0.32, 0.67], p < .001, d = 0.39).

To visualize the systematic deviation of participants' betting behavior from the prediction of Bayesian updating, we plotted the betting stakes against the models' implied win probability. Fig. 3A shows participants' betting stakes for the six pairs with equal Bayesian probability of winning. In all of these six pairs, the betting stakes were higher for low than for high SD & n trials. This visually confirms the regression results showing that risk taking systematically deviated from the predictions of the Bayesian model with known SD. For robustness, we also compared the observed betting stakes to Bayesian models assuming unknown SD and either informative or uninformative SD priors (see Supplemental Information for details). Fig. 3B and C show that these models were not able to capture the qualitative pattern of the observed betting stakes either.

For completeness, we also report the main effects of SD and sample size on betting stakes: Stakes decreased with higher SD (b = –0.034, 95 % CI [–0.041, –0.026], p < .001). Sample size had no significant influence on participants' betting stakes (b = –0.005, 95 %
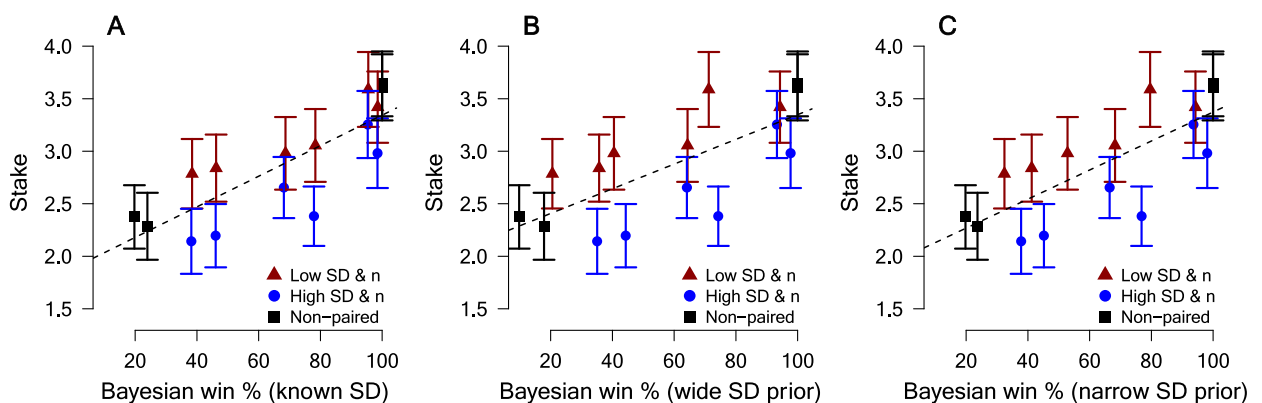

Fig. 3. *Risk taking against Bayesian measures of uncertainty in Study 1.* Betting stakes for the accuracy of the estimates for different theoretical win probabilities in Study 1. The expected Bayesian probability of winning was calculated as the probability of a sample mean being within ± 5 points of the expected value given the respective Bayesian posterior variability of the estimate for each of three models. High and low refer to the standard deviation (SD) and sample size of the sequences (high: SD = 10, 20, or 40 and n = 16 or 24; low: SD = 5, 10, or 20, and n = 4 or 6; see Table 1). Dashed lines depict linear regression lines. Error bars are 95 % confidence intervals.

CI [–0.013, 0.003], $p = .208$). These findings dovetail with the qualitative effects of SD and sample size on estimation accuracy. In summary, the results support the hypothesis that participants behaved adaptively; their betting stakes aligned with their estimation accuracy rather than with the predictions of the Bayesian benchmark models that statistically integrate all available information precisely.

3.2.3. Confidence

The observed risk-taking pattern suggests that participants have a metacognitive awareness of their own imprecision when integrating numbers. Such a metacognitive awareness should also be reflected in the confidence they have in their estimates. To test this, we elicited participants' confidence by asking them to state the probability of winning their respective bet. Results are visualized in Fig. 4A and show that participants stated higher winning probabilities in the low *SD & n* group ($M = 57.79$ %, 95 % CI [52.53, 62.06]) compared to the high *SD & n* group ($M = 46.45$ %, 95 % CI [42.13, 51.01]). This difference was significant ($b = 11.19$, 95 % CI [8.12, 14.23], $p < .001$, $d = 0.40$). This indicates that participants perceived their bets for the low *SD & n* group as less risky than for the high *SD & n* group and hence possessed a metacognitive awareness of their estimation accuracy. In line with the findings on estimation accuracy reported above, participants' perceived probability of winning decreased with higher SD ($b = –0.74$, 95 % CI [–0.86, –0.62], $p < .001$) and we found no statistically significant influence of sample size ($b = –0.12$, 95 % CI [–0.27, 0.04], $p = .144$).

Participants' actual probability of winning their bet was 51.55 % on average. A post hoc comparison of this actual probability with participants' perceived probability ($M = 52.13$ %) yields no statistically significant difference, $t(55) = 0.16$, $p > .250$, indicating that participants' confidence ratings were well calibrated on average. However, Fig. 4B hints at a regression-to-the-mean effect, in that participants overestimated winning probabilities in trials with empirical winning probabilities below 50 % ($M = 43.28$ % vs. actually 31.21 %), $t(55) = 5.35$, $p < .001$, and underestimated winning probabilities in trials with winning probabilities above 50 % ($M = 61.54$ % vs. actually 71.89 %), $t(55) = –4.24$, $p < .001$. Interestingly, four of the six trials with low *SD & n* sequences from the pairs with equal Bayesian uncertainty with known SD had winning probabilities above 50 %. Thus, although these trials had higher accuracy than the high *SD & n* sequences and participants bet higher stakes on them, these stakes might have been even higher without the regression-to-the-mean effect in participants' risk perception.

4. Study 2

So far, we implicitly assumed that participants estimated the variance of the distribution without any noise. If, however, participants' perception of the variance was imprecise or distorted as well, this could have contributed to differences between observed risk taking and the prediction of the Bayesian benchmarks with optimal information integration. To examine this, we measured how participants perceived the variances of the distributions in our task. In addition, we wanted to replicate the above results in a nonstudent online population.

4.1. Method

This study used the same set of stimuli and a similar procedure to that in Study 1 where participants repeatedly estimated the EV from a number of observed samples. In the following, we describe the differences from Study 1.
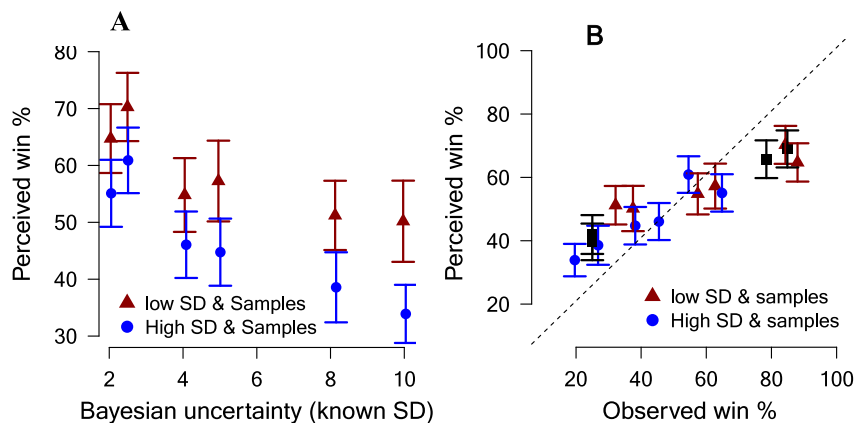

Fig. 4. *Confidence as perceived win probability for a bet in Study 1.* Perceived win percentage compared to predicted Bayesian uncertainty with known standard deviation (SD; Panel A) and compared to observed betting win rates for each combination of *SD & n* averaged across all participants (Panel B) for Study 1. In Panel A only trials from the six pairs with identical Bayesian uncertainty with known SD are depicted. In both panels high and low refer to the SD and sample size (high: *SD* = 10, 20, or 40 and *n* = 16 or 24; low: *SD* = 5, 10, or 20, and *n* = 4 or 6). Error bars are 95 % confidence intervals.

4.1.1. Study design

In the betting task, participants could choose again between five stakes, but this time the stakes ranged from £0.50 to £2.50, in steps of £0.50. Instead of the measure of confidence in Study 1, we asked participants to estimate the variability of the distributions from which they sampled as a third task. For this, we presented seven different histograms on the computer screen that visualized distributions with the same EV but different SDs: 5, 7.5, 10, 15, 20, 30, or 40. Fig. 1 displays a stylized screenshot of this task. The histograms were sorted from low to high SD and numbered from 1 (corresponding to $SD = 5$) to 7 (corresponding to $SD = 40$). In each round, participants had to select the histogram that best represented the distribution they had just experienced. To allow for an easy comparison, all plotted histograms were displayed with the same *x*- and *y*-axis range and each bar encompassed the same number range. Again, there was no feedback for any of the tasks during the experiment.

4.1.2. Procedure and incentives

Study 2 was programmed in OTree (Chen, Schonger, & Wickens, 2016) and participants were recruited online via Prolific Academic. Participants received a fixed participation fee of £2.50 and one of the 16 bets was played out for a real monetary bonus. In addition, participants could earn £0.50 if they picked the right distribution in one selected trial.

4.1.3. Participants and sample size

The number of participants was based on the results of Study 1: The effect size measured as Cohen's *d* for betting stakes between the low *SD* & *n* and the high *SD* & *n* group was 0.39. For this effect size, 105 participants gave us approximately 80 % power at a significance level of 5 % with a two-sided paired *t* test. We expected to exclude around 20 % of participants because of our predefined exclusion criteria. Therefore, we aimed to recruit 132 participants who fulfilled the following inclusion criteria: first language English, approval rate at prior Prolific tasks of at least 80 %, and a completed high school degree or equivalent. We collected informed consent from all participants and the study was approved by the Humanities and Social Sciences Research Ethics Committee at the University of Warwick.

4.1.4. Preregistration and exclusion

The study's design and its analysis were preregistered at OSF: https://osf.io/8m7fc/. Following our preregistered criteria, we excluded answers for all three dependent variables (EV estimate, betting, and variance estimate) in a trial if the EV estimate in this trial was beyond ± 5 SDs from the EV. We excluded the whole dataset of participants if they had more than five trials with EV estimates beyond ± 5 SDs, if they typed in the same answer in every trial, if they made more than two errors on the three comprehension questions, or if they stated that they did not understand their tasks at the end of the instructions. These exclusion criteria left us with 108 participants. All data, analysis code, and research materials are available at https://osf.io/36rsh/. Again, all presented regression results were from preregistered mixed-effect regressions with random participant intercepts and slopes and no interaction terms unless stated otherwise.

4.2. Results

4.2.1. Estimation accuracy and cognitive limitations

Like in Study 1, deviations from the EVs were lower in the low *SD* & *n* ($M = 6.56$, 95 % CI [5.97, 7.65]) than in the high *SD* & *n* ($M = 11.18$, 95 % CI [9.80, 14.29]) condition ($b = -0.42$, 95 % CI [-0.28, -0.55], $p < .001$, $d = -0.29$). Fig. 5A provides an overview of the observed deviations. As in Study 1, deviations from the EVs increased with higher SD ($b = 0.37$, 95 % CI [0.31, 0.44], $p < .001$),
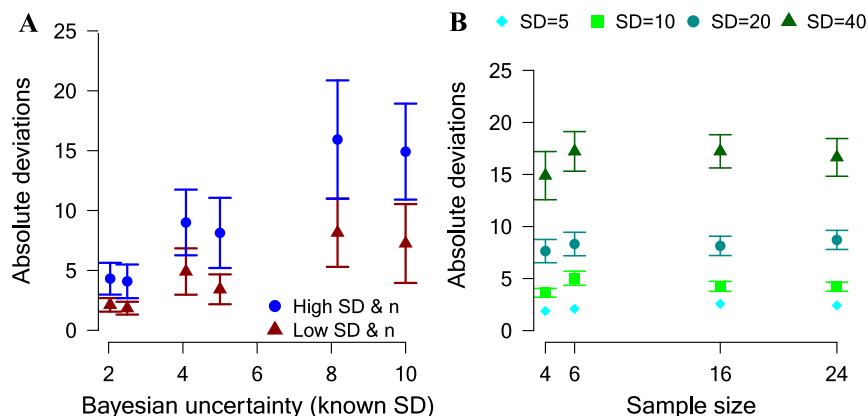

Fig. 5. *Estimation accuracy in Study 2.* Absolute deviations between estimates and expected values compared to predicted Bayesian uncertainty with known standard deviation (SD; Panel A) and compared to sample-generating SD and sample size (Panel B) for Study 2. In Panel A, only trials from the six pairs with identical Bayesian uncertainty with known SD are depicted and high and low refer to the SD and sample size (high: *SD* = 10, 20, or 40 and *n* = 16 or 24; low: *SD* = 5, 10, or 20, and *n* = 4 or 6; see Table 1). Error bars are 95 % confidence intervals.

whereas we found no evidence for an effect of sample size on participants' deviations ($b = 0.03$, 95 % CI [–0.03, 0.09], $p > .250$). The effects of SD and sample size on estimation accuracy are visualized in Fig. 5B. Finally, participants' estimates were again systematically lower than the EVs ($b = –2.70$, 95 % CI [–3.61, –1.79], $p < .001$).

4.2.2. Adaptive risk taking

Like in Study 1, participants' risk taking was aligned with the epistemic uncertainty they faced, as betting stakes were again higher in the low *SD* & *n* ($M = 2.72$, 95 % CI [2.53, 2.91]) compared to the high *SD* & *n* ($M = 2.32$, 95 % CI [2.13, 2.49]) condition ($b = 4.68$, 95 % CI [3.47, 5.90], $p < .001$, $d = 0.38$). Fig. 6 visualizes how betting stakes systematically deviated from the predictions of the Bayesian models with known (Panel A) and unknown (Panels B and C) SD that statistically integrate all available information in an optimal way. For completeness, betting stakes decreased with higher SD ($b = –0.03$, 95 % CI [–0.035, –0.025], $p < .001$) but were not significantly affected by sample size ($b = –0.004$, 95 % CI [–0.011, 0.004], $p > .250$).

4.2.3. Perception of variance

If participants' perceived variance in our task differed systematically from the true variance because of cognitive limitations in information processing, this should be considered when modeling risk taking under epistemic uncertainty. To test this, we analyzed the measures of variance perception. As can be seen from Fig. 7A, participants reliably distinguished differences in variability, as they significantly estimated higher SD sequences stemming from more variable distributions ($b = 0.07$, 95 % CI [0.063, 0.078], $p < .001$). Fig. 7A further shows that participants' variability ratings increased with sample size even though sample sizes were statistically independent of the distributional SDs in our design ($b = 0.05$, 95 % CI [0.044, 0.062], $p < .001$). Note that only the corrected sample SD is an unbiased estimated of the distributional SD, that is, subtracting one from the sample size in the denominator. In particular for small samples, the uncorrected sample SD can thus lead to judgments of variability that are too small. Thus, if participants' variability judgments are proportional to the uncorrected sample SD, this could qualitatively explain why their judgments increase with sample size. Indeed, we found empirical evidence for this explanation: A post hoc linear regression indicates that the uncorrected sample SD was a better predictor for perceived variability than the normatively suggested corrected sample SD ($dev = 6,085$ vs. $dev = 6,193$); Fig. 7B and C visualize this relationship. As participants' perception of variance was distorted, it is important to control for this distortion when modeling risk taking in the betting task, outlined next.

5. Models of risk taking under epistemic uncertainty

Participants' betting behavior in both of our two experiments was attuned to their epistemic uncertainty. This indicates a metacognitive awareness with respect to their estimation accuracy. In the following, we propose an extension of the Bayesian benchmark model of information updating outlined in the Introduction that takes this metacognitive awareness into account and hence provides a more accurate descriptive account of risk taking under epistemic uncertainty.

5.1. Mathematical specification of the Bayesian baseline model

We start our model by mathematically specifying the rules of our betting task, in which participants decided how much they wanted to bet on the accuracy of their EV estimates. The expected reward of each possible response (i.e., the betting stakes) depends on the perceived probability of winning the bet $p(w)$. For the experimental paradigm at hand with a budget of 5 CHF and the prospect of
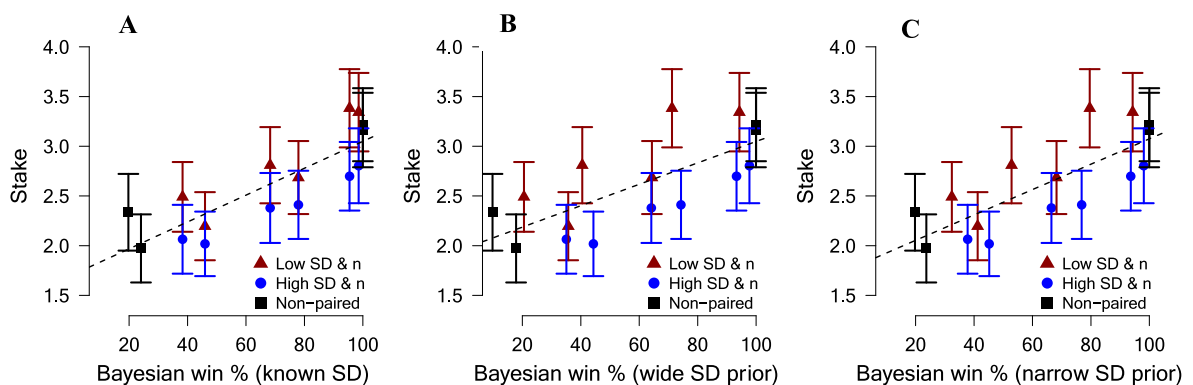

Fig. 6. *Risk taking against Bayesian measures of uncertainty in Study 2.* Betting stakes (standardized to a scale of 1 to 5 for comparison to Study 1) for the accuracy of the mean estimates for different theoretical win probabilities in Study 2. The expected Bayesian probability of winning was calculated as the probability of a sample mean being within ± 5 points of the expected value given the respective Bayesian posterior variability of the estimate for each of three models. High and low refer to the standard deviation (SD) and sample size (*n*) of the sequences (high: *SD* = 10, 20, or 40 and *n* = 16 or 24; low: *SD* = 5, 10, or 20, and *n* = 4 or 6; see Table 1). Dashed lines depict linear regression lines. Error bars are 95 % confidence intervals.
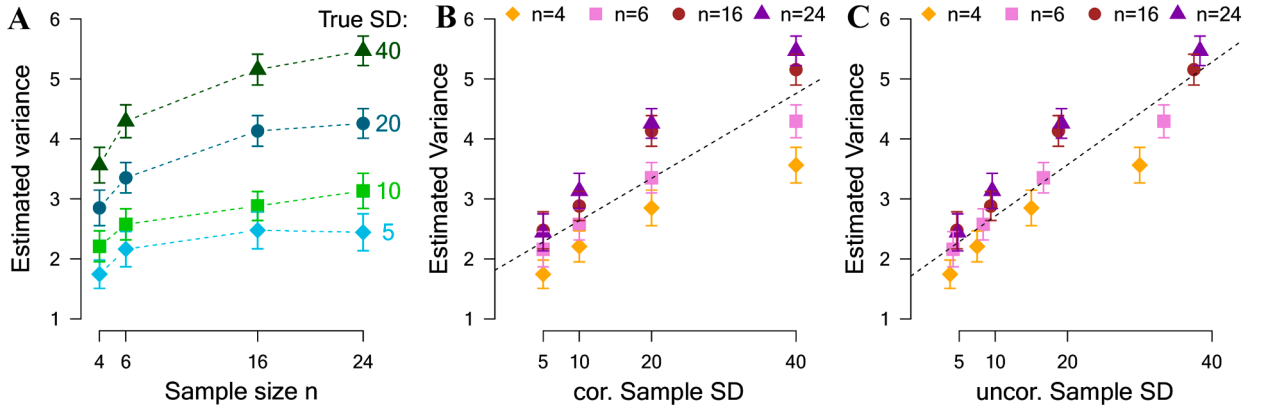
Fig. 7. *Estimates of distribution variance in Study 2.* Panel A: Estimated variance on a scale of 1 (*low variability*) to 7 (*high variability*) based on the displayed histograms dependent on standard deviation (SD) and sample size. Panel B and C: Estimated variability plotted against the corrected and uncorrected sample SD. Black dotted lines in Panels B and C are linear regression lines. All error bars are 95% confidence intervals.

doubling the stake in the case of a win, it can be calculated as follows:

$$EV(stake_j) = p(w)_j * stake_j * 2 + 5 - stake_j \tag{2}$$

with the stake restricted to integers between 1 and 5 and *j* as the index for a given trial with a specific *SD & n*. The probability of winning the bet $p(w)$ can be calculated as the probability of an estimate meeting our criterion for winning, namely, being within ± 5 points of the EV. For the normally distributed number sequences in our studies, the resulting posterior probability of winning the bet is also normally distributed. The mean of the posterior distribution is set to be equal to the EV of the sequence *M*. Thus, we assume that participants perceived their estimates as unbiased on average. The SD of the posterior distribution is calculated according to the Bayesian model with known SD as $SD_j/\sqrt{n_j}$ (see Eq. (1)). Finally, the probability of winning the bet can be calculated by subtracting the cumulative density of the posterior distribution at the point $M - 5$ from the cumulative density at the point $M + 5$:

$$p(w)_j = CDF(M_j + 5, M_j, SD_j/\sqrt{n_j}) - CDF(M_j - 5, M_j, SD_j/\sqrt{n_j}) \tag{3}$$

The SD of the posterior represents the epistemic uncertainty of winning the bet. This model assumes that participants represent the variability of their estimates according to Bayesian updating with known SD (the BUK model).

As a robustness check, we relaxed this assumption by modeling the representation of the betting uncertainty with the variability of a Bayesian updating with unknown SD model, SD_{unknown} (the BUU model). We simulated these values as described in the Supplemental Information and replaced them in the function for the probability of winning the bet as follows:

$$p(w)_j = CDF(M_j + 5, M_j, SD_{\text{unknown}}) - CDF(M_j - 5, M_j, SD_{\text{unknown}}) \tag{4}$$

The two models have in common that they take into account only the Brunswikian part of the epistemic uncertainty as represented by Bayesian updating. The only difference between the two models is the assumption about the knowledge of the true SD of the distribution.

In Study 2 we elicited participants' perceived variability of the underlying distribution. Therefore, in Study 2, we replaced the true SD in the BUK model with the perceived one of each participant in each trial $\widehat{SD}$. That way, we relax additional assumptions of these models and provide a more realistic approach, where the SD of a distribution is itself an estimate from the participants. The formula of the subjective probability of winning the bet changes to

$$p(w)_j = CDF(M_j + 5, M_j, \widehat{SD}/\sqrt{n_j}) - CDF(M_j - 5, M_j, \widehat{SD}_j/\sqrt{n_j}) \tag{5}$$

We call this Bayesian updating with subjective SD (the BUS model).

Finally, we used a stochastic link function to link the model predictions about the expected rewards of the different stakes in Eq. (2) to the observed betting stakes. For this, we assume that the stakes are multinomially distributed and that the model predictions are transformed into choice proportions according to the multinomial logit function (softmax):

$$p(stake) = \frac{\exp(\theta \cdot EV(stake))}{\sum_{k=1}^{5} \exp(\theta \cdot EV(stake_k))} \tag{6}$$

where θ governs how consistently latent reward differences between the five possible stakes, indexed with *k*, translate into choice proportions.

5.2. Descriptive models of epistemic uncertainty

If participants have a metacognitive awareness of the uncertainty they face, the descriptive accuracy of the Bayesian baseline model will be improved by augmenting it with a term that accounts for their subjective perception of the epistemic uncertainty they face. Here, we propose that this perception of epistemic uncertainty increases in proportion to the actual deviation of the participants' estimate from the EV. Since the first term, the Bayesian posterior variability, already captures the Brunswikian part of the epistemic uncertainty, this additional term captures the residual uncertainty, which we attribute to cognitive imprecisions, hence a proxy of Thurstonian uncertainty. For the BUK model this reads as follows:

$$p(w)_j = CDF\left(M_j + 5, M_j, SD_j/\sqrt{n_j} + \varphi^*\left|E_j - M_j\right|\right)$$

$$- CDF\left(M_j - 5, M_j, SD_j/\sqrt{n_j} + \varphi^*\left|E_j - M_j\right|\right) \tag{7}$$

where E is the participant's estimate and φ is a free parameter governing the individual ability to monitor one's own estimation accuracy. A φ value of zero means that a person has no awareness of their own accuracy. In contrast, a positive φ value means that the participant's deviation of the estimates from the EV increases the perceived SD of the posterior. By extension, a positive φ also means that the subjective probability of winning decreases the higher the deviation of the estimate from the EV, an adaptive response to the inaccuracy. This relation is visualized in Fig. 8A for different levels of metacognitive awareness and estimation accuracy. Note that this relation is nonlinear, meaning that there is no obviously optimal value of φ. Rather, this functional form is only an approximation of how people should take Thurstonian uncertainty into account. We call this model Bayesian updating with known SD and metacognitive awareness (BUKM). Similarly, we can augment the Bayesian updating model with the measured subjective SD in Study 2 with this metacognitive awareness term (Bayesian updating with subjective SD and metacognitive awareness or BUSM).

Besides these model specifications derived from Bayesian updating, we also examined alternative accounts of how participants in our task could represent and act upon the epistemic uncertainty they faced. The goal of this exercise was not to find the best fitting model specification but rather to examine the role of metacognitive awareness under different auxiliary modeling assumptions. First, we constructed a more flexible model, assuming that participants represent SD and sample size independently. That way the model could overweight the influence of SD compared to sample size and hence capture the additional Thurstonian uncertainty for high-SD sequences as reported in the model-free analyses. In its functional form, this model has two separate free parameters governing the impact of the SD and sample size representations on the subjective probability of winning $p(w)$.

$$p(w)_j = CDF\left(M_j + 5, M_j, \beta \cdot SD_j + \gamma \cdot \sqrt{n_j}\right) - CDF\left(M_j - 5, M_j, \beta \cdot SD_j + \gamma \cdot \sqrt{n_j}\right) \tag{8}$$

Eq. (8) depicts the case where it is assumed that the true distributional SD is known; we call this model weighted additive (WA) for *SD & n*. In Study 2 we replace the true SD again with the elicited subjective SDs (WAS). When we augment these models with a term to capture metacognitive awareness, we name these models WAM and WASM, respectively.

Second, in the models so far, we abstracted away from risk preferences. However, depending on the context, people may show signs of risk aversion, meaning that they do not choose EV maximizing but rather discount options with high variability. In our case, betting
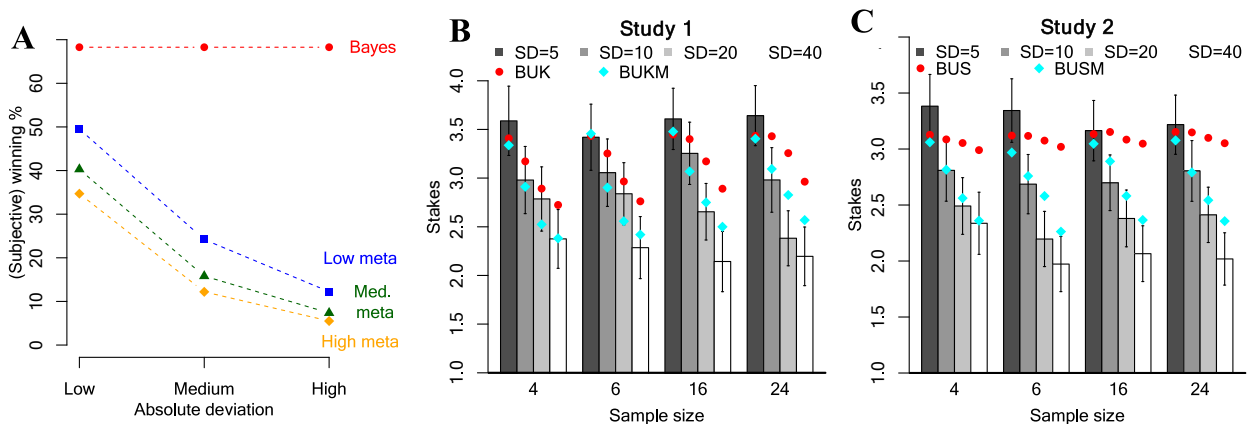

Fig. 8. *Model comparison for the effect of metacognitive awareness on risk taking in Studies 1 and 2.* Panel A: (Subjective) winning percentages for an example sequence with $SD = 20$ and sample size $n = 16$. The *x* axis shows the 25 % (low), 50 % (medium), and 75 % (high) quantile of absolute deviations of participants' estimates from the expected value. Bayes refers to the Bayesian updating model with known standard deviation (SD) and without metacognitive awareness. The other models plot the subjective winning percentages with metacognitive awareness (meta) and with the 25 % (low), 50 % (medium), and 75 % (high) quantile of hierarchically estimated individual φ values. Panels B and C: Bars are participants' average betting stakes for each combination of SD and sample size in Study 1 (B) and Study 2 (C). Dots are point predictions from hierarchical Bayesian models with metacognitive awareness (cyan) and without it (red). Error bars are 95 % confidence intervals. For the abbreviations of the models see main text. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

higher stakes increases the variance of the overall gain to our bet. Therefore, participants' betting might be influenced not only by the perceived probability of winning the bet but also by their preference for less volatile outcomes. To account for this, we implement risk preferences in an expected utility (EU) framework by transforming the possible outcomes through a power utility function with the exponent α as an additional free parameter:

$$EU\left(stake_j\right) = p(w)_j * stake_j^{\alpha} * 2 + 5 - stake_j^{\alpha} \tag{9}$$

with $\alpha < 1$ signifying risk aversion and $\alpha > 1$ risk seeking. In addition, to calculate $p(w)$ in this specification we use the BUK or BUKM model for Study 1 and the BUS or BUSM model for Study 2. When we add risk preferences to these models, we call them BUKR or BUKMR and BUSR or BUSMR, respectively.

In total we have three model types, Bayesian updating (three versions), WA, and risk preference. For all these model types we compare their respective baseline versions with the same functional form but include a term for metacognitive awareness. That way, we can check the robustness of the conclusion that metacognitive awareness plays an important part in explaining betting stakes under different assumptions about the underlying representation of uncertainty and individual differences in risk preferences.

5.3. Model estimation method

We estimated all models by means of a hierarchical Bayesian approach implemented in Stan (Carpenter et al., 2017). Using this approach, we estimated model parameters for each participant that are drawn from group-level distributions. In the following we confine ourselves to the interpretation of the parameter estimates on the group level. For all models, we defined the following priors for these group-level parameters: Choice consistency θ had a uniform group prior ranging from 0 to 3. For models that included metacognitive awareness, the group-level prior for the φ parameter was normal with mean of 0 and $SD = 10$. For the WA models and for all Bayesian updating models with metacognitive awareness, the terms for the posterior variability are not mathematically confined to positive numbers. In these cases, we forced these terms to be weakly positive. For the WA models, the weighting parameters β and γ, governing the influence of SD and sample size, respectively, both had a normal group-level prior with mean of 0 and $SD = 10$. For the risk preference models, the exponent of the power utility function α had a uniform prior between 0 and 3.

The posteriors were estimated on the basis of four independent chains with at least 1,500 samples each. Chain convergence and autocorrelation was checked for the group posteriors with the Gelman and Rubin (1992) statistic (≤ 1.02 for all reported group posteriors) and effective sample sizes (> 250 for all reported group posteriors).

5.4. Model results

5.4.1. Study 1

As our main specification, we compared the baseline Bayesian model BUK against the newly developed BUKM model, which includes metacognitive awareness. Results show that the model with metacognitive awareness predicted observed betting stakes in Study 1 better, as indicated by leave-one-out cross validation (*looic*) that takes model complexity into account (*looic* = 2,544 vs. 2,734; smaller values indicate better fit). In line with this, the metacognitive awareness term was credibly larger than zero ($\varphi = 1.97$, 95% highest probability density interval [HPDI][1.17, 2.95]). Fig. 8B shows the average risk taking across all combinations of SD and sample size in Study 1 and visualizes the superior predictive accuracy of the model with metacognitive awareness compared to the baseline model. The advantage for the BUKM model becomes visible for high *SD* & *n* trials in particular, in which the BUK model overpredicts risk taking, but the BUKM model taking Thurstonian uncertainty into account predicts lower levels of risk taking.

As robustness checks, we estimated the BUU models that relax the assumption that participants know the SD and instead assume that the SD is estimated from the observed samples. Model results show that this alternative representation of epistemic uncertainty fit the risk-taking data worse than the BUK model (*looic* = 2,738). This dovetails with the model-free results reported above showing a worse fit for risk taking based on Bayesian updating with unknown SD. Importantly, however, the fit of this model also increased when including the metacognitive awareness term (*looic* = 2,550).

We now turn to the WA model that replaces the assumption that participants represent a type of Bayesian uncertainty directly with a weighted additive representation of SD and sample size. Again, the model including the metacognitive awareness term predicted the data better than the model without the term (*looic* = 2,514 vs. 2,518). In the WAM specification with metacognitive awareness, the parameter for SD was estimated to be β = 0.59, 95%HPDI[0.40, 0.82], and for sample size *n*, γ = 0.14, 95%HPDI[−0.35, 0.52], indicating that sequences with higher SDs increase the perceived uncertainty of the estimate and thus lower the subjective probability of winning the bet. This is consistent with the regression results presented above showing that higher sequence SDs led to lower betting stakes. The insignificant effect of sample size on predicting betting stakes in this model also aligns with the results from the regression analysis.

Finally, we combined the BUK and BUKM models with an expected utility framework to incorporate risk preferences. Also in this model specification, the model including the metacognitive awareness term outperformed the model without it (*looic* = 2,456 vs. 2,706). The exponent of the utility function in the BUKMR model was α = 0.69, 95%HPDI[0.64, 0.74], corroborating risk aversion in our betting task on a group level. Importantly, even when controlling for risk preferences, the model with a metacognitive awareness term displays the better fit.

5.4.2. Study 2

Since we measured participants' perceived variance of the underlying distribution in Study 2, we could further extend the models by replacing the true SD of the distribution with the perceived one. This provides an important test for our models, as before we had to assume that participants had a veridical representation of SD. Yet, to the extent that participants actually had to estimate SD from the samples, this was an additional source of imprecision and bias that could affect risk taking. Thus, one could argue that a model with perceived SD is a more realistic representation of participants' decision process as it takes the perceived mean and the perceived SD of the observed sequences into account.

Including perceived SDs also corroborates the main finding that including the metacognitive awareness term improves model fit. For the BUSM model, φ was credibly different from zero ($\varphi = 4.18, 95\%\text{HPDI}[3.35, 5.20]$) and the *looic* was 4,866 compared to 5,430 for the BUS model without any metacognitive awareness term. Fig. 8C visualizes this result. Compared to in Study 1, the qualitative advantage in predicting risk taking for high SD sequences of the BUSM over the BUS model is enhanced. A possible reason for this advantage could be that the perceived SDs were biased compared to the true distributional SDs, as reported in the model-free analyses. Apparently, this bias in the perceived SD did not match the pattern of risk taking.

The results are similar for the WAS model with two free parameters for perceived SD and sample size. Again, the model including the term for metacognitive awareness predicted the data better (*looic* = 4,612 vs. 4,818). In the WASM model, the parameter for perceived SD was estimated to be $\beta = 0.64, 95\%\text{HPDI}[0.47, 0.84]$ and for sample size *n*, $\gamma = 0.20, 95\%\text{HPDI}[-0.22, 0.62]$. These parameter estimation results have a similar interpretation to that in Study 1, except that here, β captures the effect of perceived SD on risk taking.

Finally, also when combining the BUS and BUSM models with risk preferences, the model including the metacognitive awareness term outperformed the model without it (*looic* = 4,651 vs. 4,817). The exponent of the utility function for the BUSMR model was $\alpha = 0.81, 95\%\text{HPDI} = [0.74, 0.88]$, again showing risk aversion on a group level.

5.4.3. Modeling summary

The mean group-level posterior of the parameter indicating metacognitive awareness, φ, was credibly larger than zero in the main specifications BUKM and BUSM in both studies. This supports the idea that participants took their respective estimation accuracy into account when placing their bets. It also shows that metacognitive awareness of epistemic uncertainty is an important cognitive mechanism to understand risk taking in this task.

Besides the key result that accounting for metacognitive awareness improves model fit for all model types, a direct comparison between the models' fit metrics beyond metacognitive awareness shows that the WA model framework provides a better account of the data than the Bayesian models. The WA models have the flexibility to weight the influence of SD on risk taking independently of sample size and Bayes's theorem. Since we found an empirical link between the SD and participants' estimation accuracy, the WA models can pick up some of the regularity between SD and estimation accuracy reported in the model-free analyses. That way, the WA models cannot conceptually differentiate between a direct SD effect on risk taking and the indirect one through metacognitive awareness of the estimation accuracy that in turn affects risk taking. In the Bayesian updating models, on the other hand, the Brunswikian uncertainty represented by the Bayesian posterior variability and the Thurstonian uncertainty represented by the residual effect of the estimation accuracy on risk taking can better be distinguished. For this conceptual reason, we favor the Bayesian updating models over the WA approach, even though the latter provides a better fit to our data. Irrespective of which model one favors, our main conclusion regarding the importance of metacognitive awareness remains valid. This conclusion also holds when incorporating risk preferences into the model. Thus, taking individual differences in risk preferences into account cannot replace metacognitive awareness as a mechanism to explain risk taking in our task.

6. Discussion

In two experiments we examined people's metacognitive ability to adapt their risk-taking behavior to the epistemic uncertainty they face when sampling information. Results show that information was less accurately integrated and thus epistemic uncertainty was higher when the amount and the variability of information (i.e., sample size and SD) were high compared to low. This finding cannot be captured by Bayesian updating models that assume optimal information integration. This epistemic uncertainty was particularly high when aggregating information from highly variable samples (see also Wolfe, 1975). In addition, estimates of the distribution's EV were downward biased on average (see also Olschewski et al., 2021; Scheibehenne, 2019). This resembles performance limitations in related psychophysical and memory tasks (Cheyette & Piantadosi, 2020; Cowan, 2001; Feigenson, Dehaene, & Spelke, 2004; Stewart, Brown, & Chater, 2005). If individual decision makers cannot easily overcome these limitations, the best they can do is adapt their behavior to these constraints. In line with this, participants' risk taking was indeed adapted to their estimation accuracy when integrating information, as they took less risk in situations where their epistemic uncertainty was higher. Further, participants' subjective confidence was well aligned with their epistemic uncertainty. In summary, our results indicate that participants had metacognitive insight into their own limitations; they know that they don't know.

To model this behavior, we extended the Bayesian benchmark model of optimal information updating such that it takes metacognitive awareness into account. A model comparison confirmed that models with metacognitive awareness consistently outperformed models without it. Our modeling framework consists of two building blocks. One block is specific to our betting task as it formalizes the rules of winning the bet. The other block depicts the theoretical kernel of the model. As such, it is widely applicable to decision making under epistemic uncertainty as it describes the perceived variability of the posterior distribution of the mean estimate. This approach builds a bridge between the confidence literature and risk taking, and it applies to many real-world decisions, which

involve both aleatory and epistemic uncertainty. Hence, the experimental task at hand and the cognitive modeling approach advance our understanding of risk taking in these situations by explicitly accounting for the often-overlooked epistemic uncertainty stemming from imprecise information processing. As an example, in DfE, risk is typically assessed based on only theoretical or sampled outcome distributions (e.g., Abdellaoui, l'Haridon, & Paraschiv, 2011; Glöckner, Hilbig, Henninger, & Fiedler, 2016; Kellen, Pachur, & Hertwig, 2016; Spiliopoulos & Hertwig, 2019) and hence these analyses do not account for epistemic uncertainty when integrating information. In contrast, our approach explicitly models epistemic uncertainty as a combination of Brunswikian uncertainty, expressed as the Bayesian posterior variability of the EV estimate, and the Thurstonian part.

In situations of conscious cognitive strategy selection, metacognitive awareness is a prerequisite for adaptive behavior: When integrating sampled information in the real world, it is important to use limited cognitive resources effectively (Lieder & Griffiths, 2020; Polania, Woodford, & Ruff, 2019; Wei & Stocker, 2017; Woodford, 2020) and to choose a decision strategy adaptively (Bhui, Lai, & Gershman, 2021; Payne, Bettman, & Johnson, 1988; Vul et al., 2014). The approach at hand provides a possible channel through which this adaptation occurs: People are able to track their own epistemic uncertainty during information search and they can make effective use of this information when allocating their cognitive resources and taking risks.

In past research, people were often depicted as failing to intuitively understand the uncertainty accompanying sampled information (Tversky & Kahneman, 1971, 1974). In these studies, experimental tasks usually provided description-based statistical information. In contrast, in our experience-based task, participants adapted their risk-taking behavior to the differences in the epistemic uncertainty they faced. Thus, experiencing samples seems to foster people's intuitions about uncertainty compared to reading summarized statistical information and could be used as an educational tool to improve decision making (Bradbury, Hens, & Zeisberger, 2015; Kaufmann, Weber, & Haisley, 2013; Lejarraga & Hertwig, 2021). There are, however, many factors influencing epistemic uncertainty and we cannot rule out that other variations affecting epistemic uncertainty in our task might be harder to take into account in subsequent risk taking (e.g., changing the criterion for winning a bet, similar to the approach taken by Larrick, Burson, & Soll, 2007).

Furthermore, we found that participants' subjective perception of variance increased with larger sample sizes. This finding is consistent with participants mistaking sample for population characteristics (Fiedler, 2000; Juslin, Winman, & Hansson, 2007; Kareev et al., 2002). Consequently, risk-averse agents could favor options for which little information is available because they underestimate the variability of the reward. Unlike capacity limitations in information integration, interventions that aim to improve people's understanding of variability seem viable, for example, through extensive feedback, boosting people's statistical knowledge (Hertwig & Grüne-Yanoff, 2017), or visual aids (e.g., Goldstein & Rothschild, 2014). Irrespective of misperceptions of variability, our finding that metacognitive awareness helped participants' risk taking remain valid. We showed this by estimating the perceived variance in Study 2 and added the perceived, rather than the true, variance to our model analyses.

6.1. Comparison to related models of decisions under (partial) uncertainty

Our approach to modeling risk taking under epistemic uncertainty relates to several existing models of judgment and decision making. For example, recent economic models have picked up on the idea of noisy logarithmic representation of numeric information and have combined this with a Bayesian integration of this representation with prior information (Barretto-García, de Hollander, Grueschow, Polania, Woodford, & Ruff, 2023; Khaw et al., 2021; Petzschner, Glasauer, & Stephan, 2015). This framework connects with ours on two dimensions. First, it predicts that decision makers take the reliability of their representation relative to the prior into account when making risky choices. We add to this literature by showing that participants can indeed have a metacognitive awareness of this reliability (i.e., what we called the Thurstonian uncertainty). We differ from this literature in that we disregarded potential effects of prior information. Second, this framework relates to the idea of efficient coding, according to which the noisiness of internal representation is attuned to a given task environment, so that, for example, numeric information that is frequently encountered is represented less noisily than information that is less frequently encountered (Frydman & Jin, 2022; Polania et al., 2019; Wei & Stocker, 2015). Again, in our models we disregarded potential effects of outcome distributions on representation inaccuracies. A crucial difference to our work at hand is that these models refer to the perception of single numbers, whereas we were interested in the cognitive noise when integrating a sequence of numbers within the DfE paradigm. Thus, examining the effect of prior information and different outcome distributions on Thurstonian uncertainty and risk taking in DfE tasks depicts a promising area for future research.

Another related model class consists of two-stage models of decisions under uncertainty (Fox & Tversky, 1998; Spiliopoulos & Hertwig, 2023). In this framework, elicited probability judgments about uncertain events are used as inputs to weighting functions of probabilities under risk. This framework could also be adapted to risk taking under epistemic uncertainty by using judgments about one's own knowledge representation as input to a model of risk taking. Our models differ from this framework by using an implicit (derived from the estimation accuracy) rather than an explicit measure of epistemic uncertainty. Eventually, both approaches can incorporate the idea of metacognitive awareness in decisions under uncertainty and their usefulness should depend also on which information, implicit or stated measures of epistemic uncertainty, is accessible.

Finally, the selective integration model (Tsetsos et al., 2012, 2016) examines noisy number integration of two sequentially presented streams of numbers. Selective integration assumes that both streams compete for attention, and the weighting of numbers in the integration process depends on which stream shows the momentarily higher number. This provides an interesting avenue for extending our modeling framework of metacognitive insights to more than one option.

6.2. Metacognitive awareness and confidence

The best fitting models of risk taking under epistemic uncertainty in our data included a metacognitive awareness component.

Together with the effect of subjective confidence ratings in Study 1, this shows that metacognitive awareness is an important aspect of risk taking under epistemic uncertainty. Related to this, metacognition also plays an important role in research on subjective (over) confidence in general. Laboratory experiments in that area of research typically focus on perceptual decision making (e.g., Adler & Ma, 2018; Desender, Boldt, & Yeung, 2018; Shekhar & Rahnev, 2021). Two studies on confidence that are closer to our estimation task are by Griffin and Tversky (1992) and Kvam and Pleskac (2016). In both studies, participants had to infer the more likely of two hypotheses based on the strength of information, that is, the sample frequency for one hypothesis, and the weight of information, that is, the sample size. In both studies participants overweighted strength of evidence compared to evidence weight.

Our research extends the existing literature on (over) confidence in two ways. First, we focused on risk taking in decisions under partial uncertainty as implemented in the experimental paradigm of DfE (see also Lejarraga & Lejarraga, 2020). Hence, we implemented confidence and metacognitive awareness as constructs to build a bridge between the imprecision of numeric cognition and risk taking under epistemic uncertainty. In that way our study relates to other studies of economic decision making that explored the effect of confidence and epistemic uncertainty on binary food choices (e.g., Brus, Aebersold, Grueschow, & Polania, 2021; Folke et al., 2016). Second, most studies on confidence examined binary judgment tasks involving two options or hypotheses, whereas in our estimation task there is a potentially continuous range of hypotheses. Whereas sample size could influence confidence similarly in binary- and multiple-hypothesis evaluation tasks, the concept of strength of evidence differs between the two types of tasks. In particular, variability of information seems not to have the same influence on confidence as sample frequency in a two-hypothesis inference task (Desender et al., 2018). Therefore, from the existing confidence literature alone it did not follow that high variability increased task difficulty. This is because from a Brunswikian perspective, the difficulty of estimating the EV depends on the SD of the posterior distribution, in which information variability is only one part. Our contribution is novel because we filled this gap and empirically showed that information variability increased integration difficulty as a result of an increase in Thurstonian uncertainty and further showed that people were aware of this in their risk taking.

6.3. Limitations

In Study 1, our participants came from a student population, whereas in Study 2, we recruited from a general English-speaking population online. We therefore believe that our results are generalizable to the general population, although we presume a minimal understanding of numeric information is necessary to have a metacognitive awareness of one's own cognitive limitations in the processing of numeric information. This suggests that there are individual differences in metacognitive awareness and how it affects risk taking. From this perspective, it seems plausible to predict that participants with a stronger metacognitive awareness (in our model a higher φ) would earn more money in our betting task because they can better assess when they should bet high or low stakes. However, post hoc we found no significant relation between total money won in the bets and the individual metacognitive awareness parameter estimates in our model. This could be due to the relatively low number of measurement repetitions per participant and hence relatively low statistical power to analyze individual differences. There was also a strong correlation between the number of bets won and the money won, indicating that individual differences in monetary success were mainly determined by estimation accuracy rather than metacognitive awareness. Metacognitive awareness played a crucial role in understanding betting-stake differences between sequences with different SD & n within participants, though. Future research could examine individual differences in metacognitive awareness and risk taking in a more tailored design. Another promising avenue for future research would be to include individual difference measures such as numeracy or symbolic number mapping (e.g., Cokely, Galesic, Schulz, Ghazal, & Garcia-Retamero, 2012; Schley & Peters, 2014). Similarly, an independent assessment of individual risk preferences could help elucidate the selective impact of preferences and metacognitive awareness on risk taking in more detail (but see Frey, Pedroni, Mata, Rieskamp, & Hertwig, 2017, for the difficulty of measuring risk preferences across tasks).

To alleviate the concern that results were driven by participants who misunderstood the task by betting on estimates of the sample mean rather than the distribution's EV, we conducted an additional study in which participants estimated the sample mean (see the Supplementary Information for details). Results from this study reveal that participants' risk taking differed from patterns observed in the two main studies presented here. In particular, risk taking was negatively affected by the sample size in the study where participants' task was to estimate the sample mean, consistent with the idea that participants understood that more samples might be integrated with more noise. However, this was not the case when participants estimated the EV, as reported in the two main studies of this article.

6.4. Conclusion

To summarize, our analyses suggest that the understanding of human judgment and decision making benefits from explicitly modeling people's epistemic uncertainty and their metacognitive awareness thereof. Quantifying uncertainty solely on the basis of statistical models such as Bayesian posterior variability ignores limitations in the cognitive processing of numeric information that can (and should) affect behavior (Khaw et al., 2021; Schley & Peters, 2014). Furthermore, an explicit model of epistemic uncertainty can contribute to a better understanding of human information search and retrieval, as well as to strategy selection (Gershman, 2019; Hertwig & Pleskac, 2010; Lejarraga, Hertwig, & Gonzalez, 2012; Lieder, Griffiths, & Hsu, 2018; Pirolli & Card, 1999; Stewart, Chater, & Brown, 2006; Vul et al., 2014; Zhu, Sanborn, & Chater, 2020). In a broader context, taking the cognitive processes of information perception and integration into account helps explain and predict behavior under partial uncertainty, such as in DfE (Mason, Madan, Simonsen, Spetch, & Ludwig, 2022; Olschewski, Luckman, Mason, Ludwig, & Konstantinidis, 2024). Modeling judgment and decision-making processes at the interplay between inevitable cognitive constraints and epistemic uncertainty is key to understanding human

behavior in a fundamentally uncertain world. If people know that they don't know, what looks like irrational behavior at first glance may turn out to be a sensible adaptation.

Funding

CRediT authorship contribution statement

Sebastian Olschewski: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Benjamin Scheibehenne:** Conceptualization, Funding acquisition, Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data is available on osf and the link is shared in the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cogpsych.2024.101642.

References

Abdellaoui, M., l'Haridon, O., & Paraschiv, C. (2011). Experienced vs. described uncertainty: Do we need two prospect theory specifications? *Management Science, 57*(10), 1879–1895.

Adler, W. T., & Ma, W. J. (2018). Comparing Bayesian and non-Bayesian accounts of human confidence reports. *PLoS Computational Biology, 14*(11), Article e1006572.

Barretto-García, M., de Hollander, G., Grueschow, M., Polania, R., Woodford, M., & Ruff, C. C. (2023). Individual risk attitudes arise from noise in neurocognitive magnitude representations. *Nature Human Behaviour, 7*(9), 1551–1567.

Barron, G., & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making, 16*(3), 215–233. https://doi.org/10.1002/bdm.443

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Beach, L. R., & Swenson, R. G. (1966). Intuitive estimation of means. Psychonomic. *Science, 5*(4), 161–162.

Benjamin, D. J. (2019). Errors in probabilistic reasoning and judgment biases. *Handbook of Behavioral Economics: Applications and Foundations, 1*(2), 69–186.

Bhui, R., Lai, L., & Gershman, S. J. (2021). Resource-rational decision making. *Current Opinion in Behavioral Sciences, 41*, 15–21.

Boundy-Singer, Z. M., Ziemba, C. M., & Goris, R. L. (2023). Confidence reflects a noisy decision reliability estimate. *Nature Human Behaviour, 7*(1), 142–154.

Bradbury, M. A., Hens, T., & Zeisberger, S. (2015). Improving investment decisions with simulated experience. *Review of Finance, 19*(3), 1019–1052.

Brezis, N., Bronfman, Z. Z., & Usher, M. (2015). Adaptive spontaneous transitions between two mechanisms of numerical averaging. *Scientific Reports, 5*(1), 10415.

Brus, J., Aebersold, H., Grueschow, M., & Polania, R. (2021). Sources of confidence in value-based choice. *Nature Communications, 12*(1), 7337.

Budescu, D. V., & Du, N. (2007). Coherence and consistency of investors' probability judgments. *Management Science, 53*(11), 1731–1744.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, 76*(1), 1–32. https://doi.org/10.18637/jss.v076.i01

Charness, G., & Levin, D. (2005). When optimal choices feel wrong: A laboratory study of Bayesian updating, complexity, and affect. *American Economic Review, 95*(4), 1300–1309. https://doi.org/10.1257/0002828054825583

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance, 9*, 88–97.

Cheyette, S. J., & Piantadosi, S. T. (2020). A unified account of numerosity perception. *Nature Human Behaviour, 4*(12), 1265–1272.

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin numeracy test. *Judgment and Decision Making, 7*(1), 25–47. https://doi.org/10.1017/S1930297500001819

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Dall, S. R., Giraldeau, L. A., Olsson, O., McNamara, J. M., & Stephens, D. W. (2005). Information and its use by animals in evolutionary ecology. *Trends in Ecology & Evolution, 20*(4), 187–193. https://doi.org/10.1016/j.tree.2005.01.010

Dasgupta, I., Schulz, E., Tenenbaum, J. B., & Gershman, S. J. (2020). A theory of learning to infer. *Psychological Review, 127*(3), 412–441.

Desender, K., Boldt, A., & Yeung, N. (2018). Subjective confidence predicts information seeking in decision making. *Psychological Science, 29*(5), 761–778.

Enke, B., & Graeber, T. (2023). Cognitive uncertainty. *The Quarterly Journal of Economics, 138*(4), 2021–2067.

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over-and underconfidence: The role of error in judgment processes. *Psychological Review, 101*(3), 519–527.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307–314.

Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review, 107*(4), 659–676. https://doi.org/10.1037/0033-295X.107.4.659

Folke, T., Jacobsen, C., Fleming, S. M., & De Martino, B. (2016). Explicit representation of confidence informs future value-based decisions. *Nature Human Behaviour, 1*(1), 0002.

Fox, C. R., & Tversky, A. (1998). A belief-based account of decision under uncertainty. *Management Science, 44*(7), 879–895.

Fox, C. R., & Ülkümen, G. (2011). Distinguishing two dimensions of uncertainty. In W. Brun, G. Keren, G. Kirkebøen, & H. Montgomery (Eds.), *Perspectives on thinking, judging, and decision making* (pp. 21–35). Scandinavian University Press.

Frey, R., Pedroni, A., Mata, R., Rieskamp, J., & Hertwig, R. (2017). Risk preference shares the psychometric structure of major psychological traits. *Science Advances, 3*(10), Article e1701381.

Frydman, C., & Jin, L. J. (2022). Efficient coding and risky choice. *Quarterly Journal of Economics, 137*(1), 161–213.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis.* Chapman and Hall/CRC.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*(4), 457–472.

Gershman, S. J. (2019). Uncertainty and exploration. *Decision, 6*(3), 277–286. https://doi.org/10.1037/dec0000101

Glöckner, A., Hilbig, B. E., Henninger, F., & Fiedler, S. (2016). The reversed description-experience gap: Disentangling sources of presentation format effects in risky choice. *Journal of Experimental Psychology: General, 145*(4), 486–508.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24*(3), 411–435.

Goldstein, D. G., & Rothschild, D. (2014). Lay understanding of probability distributions. *Judgment & Decision Making, 9*(1), 1–14. https://doi.org/10.1017/S1930297500004940

Hacking, I. (2006). *The emergence of probability: A philosophical study of early ideas about probability, induction and statistical inference.* Cambridge University Press.

Hahn, U. (2014). Experiential limitation in judgment and decision. *Topics in Cognitive Science, 6*(2), 229–244. https://doi.org/10.1111/tops.12083

Heath, C., & Tversky, A. (1991). Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of Risk and Uncertainty, 4*(1), 5–28.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science, 15*(8), 534–539. https://doi.org/10.1111/j.0956-7976.2004.00715.x

Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in Cognitive Sciences, 13*(12), 517–523.

Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science, 12*(6), 973–986.

Hertwig, R., & Pleskac, T. J. (2010). Decisions from experience: Why small samples? *Cognition, 115*(2), 225–237. https://doi.org/10.1016/j.cognition.2009.12.009

Holt, C. A., & Smith, A. M. (2009). An update on Bayesian updating. *Journal of Economic Behavior & Organization, 69*(2), 125–134.

Johnson, D. D., & Fowler, J. H. (2011). The evolution of overconfidence. *Nature, 477*(7364), 317–320.

Juslin, P., & Olsson, H. (1997). Thurstonian and Brunswikian origins of uncertainty in judgment: A sampling model of confidence in sensory discrimination. *Psychological Review, 104*(2), 344–366. https://doi.org/10.1037/0033-295X.104.2.344

Juslin, P., Winman, A., & Hansson, P. (2007). The naïve intuitive statistician: A naïve sampling model of intuitive confidence intervals. *Psychological Review, 114*(3), 678–703. https://doi.org/10.1037/0033-295X.114.3.678

Kareev, Y., Arnon, S., & Horwitz-Zeliger, R. (2002). On the misperception of variability. *Journal of Experimental Psychology: General, 131*(2), 287–297. https://doi.org/10.1037/0096-3445.131.2.287

Kaufmann, C., Weber, M., & Haisley, E. (2013). The role of experience sampling and graphical displays on one's investment risk appetite. *Management Science, 59*(2), 323–340. https://doi.org/10.1287/mnsc.1120.1607

Kellen, D., Pachur, T., & Hertwig, R. (2016). How (in) variant are subjective representations of described and experienced risk and rewards? *Cognition, 157*, 126–138.

Khaw, M. W., Li, Z., & Woodford, M. (2021). Cognitive imprecision and small-stakes risk aversion. *The Review of Economic Studies, 88*(4), 1979–2013.

Konovalova, E., & Le Mens, G. (2020). An information sampling explanation for the in-group heterogeneity effect. *Psychological Review, 127*(1), 47–73. https://doi.org/10.1037/rev0000160

Kuhnen, C. M. (2015). Asymmetric learning from financial information. *The Journal of Finance, 70*(5), 2029–2062.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(1), 1–26.

Kvam, P. D., & Pleskac, T. J. (2016). Strength and weight: The determinants of choice and confidence. *Cognition, 152*, 170–180.

Laestadius, J. E., Jr. (1970). Tolerance for errors in intuitive mean estimations. *Organizational Behavior and Human Performance, 5*(2), 121–124.

Lee, D. G., & Usher, M. (2023). Value certainty in drift-diffusion models of preferential choice. *Psychological Review, 130*(3), 790.

Larrick, R. P., Burson, K. A., & Soll, J. B. (2007). Social comparison and confidence: When thinking you're better than average predicts overconfidence (and when it does not). *Organizational Behavior and Human Decision Processes, 102*(1), 76–94.

Lebreton, M., Abitbol, R., Daunizeau, J., & Pessiglione, M. (2015). Automatic integration of confidence in the brain valuation signal. *Nature Neuroscience, 18*(8), 1159–1167.

Lejarraga, T., & Hertwig, R. (2021). How experimental methods shaped views on human competence and rationality. *Psychological Bulletin, 147*(6), 535–564. https://doi.org/10.1037/bul0000324

Lejarraga, T., Hertwig, R., & Gonzalez, C. (2012). How choice ecology influences search in decisions from experience. *Cognition, 124*(3), 334–342.

Lejarraga, T., & Lejarraga, J. (2020). Confidence and the description–experience distinction. *Organizational Behavior and Human Decision Processes, 161*, 201–212.

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences, 43*. https://doi.org/10.1017/S0140525X1900061X

Lieder, F., Griffiths, T. L., & Hsu, M. (2018). Overrepresentation of extreme events in decision making reflects rational use of cognitive resources. *Psychological Review, 125*(1), 1–32. https://doi.org/10.1037/rev0000074

Ludvig, E. A., & Spetch, M. L. (2011). Of black swans and tossed coins: is the description-experience gap in risky choice limited to rare events? *PloS One, 6*(6), Article e20262.

Malmendier, U., & Tate, G. (2015). Behavioral CEOs: The role of managerial overconfidence. *Journal of Economic Perspectives, 29*(4), 37–60.

March, J. G. (2010). *The ambiguities of experience.* Cornell University Press.

Mason, A., Madan, C. R., Simonsen, N., Spetch, M. L., & Ludvig, E. A. (2022). Biased confabulation in risky choice. *Cognition, 229*, Article 105245.

Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Confidence as Bayesian probability: From neural origins to behavior. *Neuron, 88*(1), 78–92.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81–97. https://doi.org/10.1037/h0043158

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115*(2), 502–517. https://doi.org/10.1037/0033-295X.115.2.502

Navajas, J., Hindocha, C., Foda, H., Keramati, M., Latham, P. E., & Bahrami, B. (2017). The idiosyncratic nature of confidence. *Nature Human Behaviour, 1*(11), 810–818.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101*(4), 608–631. https://doi.org/10.1037/0033-295X.101.4.608

Obrecht, N. A., Chapman, G. B., & Gelman, R. (2007). Intuitive t tests: Lay use of statistical information. *Psychonomic Bulletin & Review, 14*(6), 1147–1152. https://doi.org/10.3758/BF03193104

Olschewski, S., Luckman, A., Mason, A., Ludvig, E., & Konstantinidis, E. (2024). The future of decisions from experience: Connecting real-world decision problems to cognitive processes. *Perspectives on Psychological Science, 19*(1), 82–102. https://doi.org/10.1177/17456916231179138

Olschewski, S., Newell, B. R., Oberholzer, Y., & Scheibehenne, B. (2021). Valuation and estimation from experience. *Journal of Behavioral Decision Making, 34*(5), 729–741.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 534–552. https://doi.org/10.1037/0278-7393.14.3.534

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics, 2.* https://doi.org/10.3389/neuro.11.010.2008

Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin, 68*(1), 29–46. https://doi.org/10.1037/h0024722

Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences, 19*(5), 285–293.

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review, 106*(4), 643–675. https://doi.org/10.1037/0033-295X.106.4.643

Polania, R., Woodford, M., & Ruff, C. C. (2019). Efficient coding of subjective value. *Nature Neuroscience, 22*(1), 134–142.

R Core Team. (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/.

Rosenbaum, D., Glickman, M., Fleming, S. M., & Usher, M. (2022). The cognition/metacognition trade-off. *Psychological Science, 33*(4), 613–628.

Scheibehenne, B. (2019). The psychophysics of number integration: Evidence from the lab and from the field. *Decision, 6*(1), 61–76. https://doi.org/10.1037/dec0000089

Schley, D. R., & Peters, E. (2014). Assessing "economic value" symbolic-number mappings predict risky and riskless valuations. *Psychological Science, 25*(3), 753–761.

Shekhar, M., & Rahnev, D. (2021). Sources of metacognitive inefficiency. *Trends in Cognitive Sciences, 25*(1), 12–23.

Simon, H. A. (1959). Theories of decision-making in economics and behavioral science. *American Economic Review, 49*(3), 253–283.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics, 69*(1), 99–118.

Simon, H. A. (1986). Rationality in psychology and economics. *Journal of Business, 59*, S209–S224.

Soll, J. B., Palley, A., Klayman, J., & Moore, D. (in press). Overconfidence in probability distributions: People know they don't know but they don't know what to do about it. *Management Science*.

Spencer, J. (1963). A further study of estimating averages. *Ergonomics, 6*(3), 255–265.

Spiliopoulos, L., & Hertwig, R. (2019). Nonlinear decision weights or moment-based preferences? A model competition involving described and experienced skewness. *Cognition, 183*, 99–123.

Spiliopoulos, L., & Hertwig, R. (2023). Variance, skewness and multiple outcomes in described and experienced prospects: Can one descriptive model capture it all? *Journal of Experimental Psychology: General, 152*(4), 1188–1222. https://doi.org/10.1037/xge0001323

Spitzer, B., Waschke, L., & Summerfield, C. (2017). Selective overweighting of larger magnitudes during noisy numerical comparison. *Nature Human Behaviour, 1*(8), 0145. https://doi.org/10.1038/s41562-017-0145

Stewart, N., Brown, G. D., & Chater, N. (2005). Absolute identification by relative judgment. *Psychological Review, 112*(4), 881–911. https://doi.org/10.1037/0033-295X.112.4.881

Stewart, N., Chater, N., & Brown, G. D. (2006). Decision by sampling. *Cognitive Psychology, 53*(1), 1–26.

Tannenbaum, D., Fox, C. R., & Ülkümen, G. (2017). Judgment extremity and accuracy under epistemic vs. aleatory uncertainty. *Management Science, 63*(2), 497–518.

Tauber, S., Navarro, D. J., Perfors, A., & Steyvers, M. (2017). Bayesian models of cognition revisited: Setting optimality aside and letting data drive psychological theory. *Psychological Review, 124*(4), 410–441. https://doi.org/10.1037/rev0000052

Tsetsos, K., Chater, N., & Usher, M. (2012). Salience driven value integration explains decision biases and preference reversal. *Proceedings of the National Academy of Sciences of the United States of America, 109*(24), 9659–9664.

Tsetsos, K., Moran, R., Moreland, J., Chater, N., Usher, M., & Summerfield, C. (2016). Economic irrationality is optimal during noisy decision making. *Proceedings of the National Academy of Sciences of the United States of America, 113*(11), 3102–3107.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*(2), 105–110. https://doi.org/10.1037/h0031322

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? Optimal decisions from very few samples. *Cognitive Science, 38*(4), 599–637. https://doi.org/10.1111/cogs.12101

Walters, D. J., Fernbach, P. M., Fox, C. R., & Sloman, S. A. (2017). Known unknowns: A critical determinant of confidence and calibration. *Management Science, 63*(12), 4298–4307.

Walters, D. J., Ülkümen, G., Tannenbaum, D., Erner, C., & Fox, C. R. (2023). Investor Behavior Under Epistemic vs. *Aleatory Uncertainty. Management Science, 69*(5), 2761–2777.

Wei, X. X., & Stocker, A. A. (2015). A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts. *Nature Neuroscience, 18*(10), 1509–1517.

Wei, X. X., & Stocker, A. A. (2017). Lawful relation between perceptual bias and discriminability. *Proceedings of the National Academy of Sciences of the United States of America, 114*(38), 10244–10249.

Wolfe, M. L. (1975). Distribution characteristics as predictors of error in intuitive estimation of means. *Psychological Reports, 36*(2), 367–370.

Woodford, M. (2020). Modeling imprecision in perception, valuation, and choice. *Annual Review of Economics, 12*, 579–601.

Wulff, D. U., Mergenthaler-Canseco, M., & Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin, 144*(2), 140–176. https://doi.org/10.1037/bul0000115

Xiang, Y., Graeber, T., Enke, B., & Gershman, S. J. (2021). Confidence and central tendency in perceptual judgment. *Attention, Perception, & Psychophysics*, 1–11.

Zhu, J. Q., Sanborn, A. N., & Chater, N. (2020). The Bayesian sampler: Generic Bayesian inference causes incoherence in human probability judgments. *Psychological Review, 127*(5), 719–748. https://doi.org/10.1037/rev0000190