


RESEARCH ARTICLE

WILEY

Valuation and estimation from experience

Sebastian Olschewski^{1,2}  | Ben R. Newell³ | Yvonne Oberholzer⁴ | Benjamin Scheibehenne⁴

¹Warwick Business School, University of Warwick, Coventry, UK

²Department of Psychology, University of Basel, Basel, Switzerland

³School of Psychology, University of New South Wales, Sydney, New South Wales, Australia

⁴Institute of Information Systems and Marketing, Karlsruhe Institute of Technology, Karlsruhe, Germany

Correspondence

Sebastian Olschewski, Warwick Business School, University of Warwick, Coventry, CV4 7AL, UK.

Email: Sebastian.olschewski@wbs.ac.uk

Funding information

Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: 100014_149846, 100014_175849 and P2BSP1_188188

Abstract

The processing of sequentially presented numerical information is a prerequisite for decisions from experience, where people learn about potential outcomes and their associated probabilities and then make choices between gambles. Little is known, however, about how people's preference for choosing a gamble is affected by how they perceive and process numerical information. To address this, we conducted a series of experiments wherein participants repeatedly sampled numbers from continuous outcome distributions. They were incentivized either to estimate the means of the numbers or to state their minimum selling prices to forgo a consequential draw from the distributions (i.e., the certainty equivalents or valuations). We found that participants valued distributions below their means, valued high-variance sequences lower than low-variance sequences, and valued left-skewed sequences lower than right-skewed sequences. Though less pronounced, similar patterns occurred in the mean estimation task where preferences should not play a role. These results are not consistent with prior findings in decision from experience such as the overweighting of high numbers and the underweighting of rare events. Rather, the qualitative effects, as well as the similarity of effects in valuation and estimation, are consistent with the assumption that people process numbers on a compressed mental number line in valuations from experience.

KEYWORDS

continuous outcome distributions, decision from experience, mean estimation, mental number line, numeric cognition, risk preference, valuation

1 | INTRODUCTION

In many real-world situations, people experience decision outcomes over time. In an economic context, decision makers must integrate numerical information such as returns on investments that accrue sequentially. In experimental research, this scenario is modeled using the decisions from experience (DFEs) paradigm (Hertwig et al. 2004; Barron & Erev, 2003; Weber et al. 2004), where people can freely sample from an outcome distribution and afterward make a consequential choice. This contrasts with the decisions from description (DFD) paradigm, where all relevant information is readily summarized.

Recently, there has been debate about whether economic behavior in DFD can be explained partly by regularities in the perception of numeric information (Schley & Peters, 2014; Bordalo et al. 2012; Woodford, 2020). Yet the extent to which this could also be the case in DFE, where understanding a choice option requires the perception, memorization, and integration of sequential numeric information, is an open question.

Prior research on DFE comprised mostly studies with choices between one- or two-outcome gambles. In this study, we broaden the scope to subjective valuations (or certainty equivalents) of continuous number streams and examine how they compare with choice behavior

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. Journal of Behavioral Decision Making published by John Wiley & Sons Ltd.

reported in DFE. In addition, we compare these valuations to objective mean judgments to examine whether behavioral patterns extend to nonpreferential tasks. Whereas preferential valuations are usually subject to risk and skewness preferences, mean estimates should not be affected by these preferences. To the extent that numeric information is similarly processed in both objective estimation and subjective valuation tasks, behavioral patterns should be similar in both tasks. However, to the extent that risk and skewness preferences drive subjective valuations but not objective estimations, responses in valuation should differ from those in the estimation task. In the following, we develop specific hypotheses that are rooted in existing findings in the literature.

1.1 | Underweighting of rare events and overweighting of high outcomes

Past research in DFE suggests that people choose as if they underweight rare events (Hertwig et al. 2004). Although this effect may be partly due to undersampling of the rare event (Fox & Hadar, 2006), a recent meta-analysis found that the effect remains when controlling for undersampling (Wulff et al. 2018). Another important finding is that people are risk seeking in DFE, which suggests that participants overweight high outcomes (Ludvig & Spetch, 2011). This overweighting has been corroborated by higher frequency judgments for higher numbers in the gain domain (Madan et al. 2014, 2016).

So far, underweighting of rare events and overweighting of high outcomes have been tested predominantly on choices for gambles with one or two discrete outcomes. Using valuations of continuous outcome sequences, however, allows us to test the generalizability of these effects and provides a bridge to the numeric cognition literature. For symmetric distributions like the normal distribution, rare events are equally likely to occur for high and low outcomes. Thus, underweighting of rare events predicts no valuations below the mean and no effect of the variance on valuations. In skewed distributions, rare events are more likely to occur on one side of the distribution. Hence, if all other characteristics are equal, underweighting of rare events predicts lower valuations for right-skewed distributions (where high values are rare) than for left-skewed distributions (where low values are rare). In contrast, overweighting of high outcomes predicts higher valuations than the mean, and higher valuations for high-variance (than for low variance) and right-skewed (than for left skewed) distributions, again if all other characteristics are equal.

1.2 | Numeric cognition and the compressed mental number line

Important contributions have been made to the better understanding of how people process and integrate information sequentially (e.g., Ashby & Rakow, 2014; Baucells et al. 2011; Hotaling et al. 2019; Wulff & Pachur, 2016). However, most of this research has not connected to number perception research directly. To bridge this gap, we develop a hypothesis for valuations from experience from the perspective of numeric cognition.

People have an inherently imprecise and nonverbal notion of numerosity (Cheyette & Piantadosi, 2019; Gallistel & Gelman, 2000; Moyer & Landauer, 1967; Whalen et al. 1999). Research by Dehaene

and colleagues (Dehaene, 2003; Dehaene et al. 2008; Feigenson et al. 2004) has indicated that the internal representation of numerals can be described as a compressed mental number line. This means that differences in numerosity are represented to be smaller for higher numerosity, similar to concave mappings from objective stimuli to subjective perceptions in other domains (see Fechner, 1860).

For the processing of number sequences, the compressed mental number line predicts that people give lower estimations, and hence also lower valuations, than the true mean of the underlying distribution. Further, it predicts that both estimation and valuation are lower for sequences with higher variance. In this case, apparent risk-averse behavior would be partly due to numeric perception rather than subjective economic preference (Schoemaker, 1982). Furthermore, it can be shown mathematically that for compressed power functions, sequence means are higher for right- than for left-skewed distributions (Genest et al. 2016; Menezes et al. 1980). This implies that people also value right-skewed distributions higher than left-skewed ones and that seeming skewness preferences could be rooted partly in the perception of numbers.

The empirical evidence for a compressed mental number line when integrating symbolic number sequences is mixed. Whereas some studies have reported that people have a tendency to underestimate the mean or the sum of a number sequence (Brezis et al. 2015; Scheibehenne, 2019), others have found no such effect (e.g., Lindskog & Winman, 2014; Peterson & Beach, 1967). Consequently, it is an open question if and under what circumstances a compressed mental number line applies when integrating symbolic continuous number sequences.

1.3 | Summary of hypotheses

All other characteristics being equal, underweighting of rare events predicts an economic preference for left- over right-skewed distributions and risk neutrality for symmetrically distributed gambles. In contrast, overweighting of high numbers predicts a preference for right-skewed over left-skewed gambles and risk seeking for gains. Finally, the compressed mental number line predicts overall risk aversion and a preference for right-skewed over left-skewed gambles. These predictions are summarized in Table 1.

To systematically test these predictions, we conducted a series of experiments in which participants repeatedly sampled outcomes from continuous payoff distributions that differed in their mean, variance, and skewness. For each distribution, participants were incentivized either to make an economic valuation or to estimate the mean.

2 | EXPERIMENT 1A

2.1 | Method

2.1.1 | The tasks

The experiment consisted of economic valuation and numeric estimation tasks in a within-subjects design. Before giving their answer, participants could freely sample outcomes from an underlying distribution. In the valuation task, participants were asked to state their minimum selling prices (willingness-to-accept) or certainty equivalent

	Underweighting of rare events	Overweighting of high numbers	Compressed mental number line
Symmetric distribution	0	+	-
Effect of variance	0	+	-
Effect of skewness	-	+	+

Note: Behavioral predictions based on findings in decision from experience (DFE) and numeric cognition on the valuation of a continuous number sequence. Key: 0 means no deviation from the mean or no effect, + means a positive deviation from the mean or a positive effect, and - means a negative deviation from the mean or a negative effect. For details, see text.

for each distribution. This was explained to them as the minimum price they would demand to forgo the option to make a single consequential draw from that distribution that would be paid out. In the estimation task, participants were asked to estimate the mean of the distribution. Here, accuracy was incentivized with respect to how closely the estimates matched the theoretical mean of the underlying distribution.

Under the assumption that participants were well calibrated to the mean estimation task, the variability of the monetary bonus should have been higher in the valuation than in the estimation task. However, we believe that this is at the core of the difference between the two tasks and that other incentives of the estimation task would have made answers subject to risk preferences.

Each trial contained a rectangular box representing the underlying outcome distribution. Participants could sample from the distribution by pressing <space>. Each sampled number was shown for 250 ms and was generated as a random draw from the respective underlying distribution, rounded to its nearest integer. Participants had to draw at least one sample before they typed their answers into the gray fields and confirmed their inputs with <enter> (see Figure 1 for a schematic).

2.1.2 | Distributions

We constructed 24 continuous number distributions by combining four means (80, 100, 130, and 160), two standard deviations (5 and 10), and three distribution shapes (normal, left skewed, and right skewed). Skewed distributions were constructed from scaled gamma distributions with a shape parameter of 1 (absolute skewness = 2) and were truncated at the first (left skewed) or last (right skewed) percentile to avoid extreme outliers. The different distributions were presented in randomized order and were the same in both the valuation and the estimation tasks.

2.1.3 | Procedure and incentives

The experiment was implemented on a computer with PsychoPy (Peirce, 2007) and conducted in individual sessions in separate rooms at the University of New South Wales School of Psychology. All instructions were presented on the computer screen and could be read at a participant's own pace. Each participant completed two blocks of 24 tasks, starting with either estimations or valuations.

Payment was determined by randomly selecting one answer across both blocks. If the trial was in the valuation block, a Becker-DeGroot-Marschak (BDM) procedure was implemented (Becker et al. 1964): a random number was uniformly drawn between zero and the theoretical mean of a given distribution. When the random number was below the participant's answer for this trial, the

TABLE 1 Predictions for valuations from experience

participant received a draw from the distribution; otherwise, the participant received the points from the random number for certain. If the selected trial was in the estimation block, the observed points were determined by the true mean from which the error of the estimate was subtracted. Finally, the obtained points were exchanged for Australian dollars (AUD) with a 20:1 ratio and paid out in cash.

2.1.4 | Participants and data analysis

We tested 53 participants and determined sample size prior to data inspection. Participants were undergraduates from the University's subject pool, recruited via online advertisement. Participants received course credit and a choice-dependent bonus of 1.50 to 8.93 AUD ($M_{\text{pay}} = 5.43$ AUD). In the subject pool, the mean age was 19 years, and approximately 70% of the subjects were women.

Prior to analyzing the data, we excluded two participants who did not comply with the task. Further, we excluded answers more than five standard deviations from the distribution's mean (21 out of 2448 total trials). We assumed that in these trials, participants made typos or did not pay attention to the samples; thus, these trials were not informative for our research question.

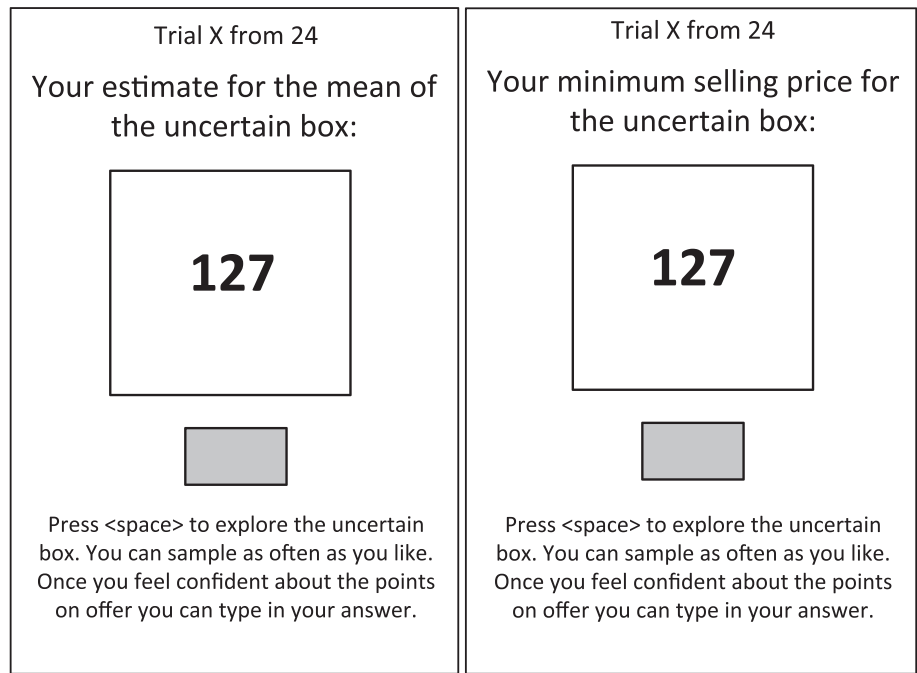
We analyzed the data by means of a participant mixed-effects regression analysis in R (R Core Team, 2016; RStudio Team, 2015) using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) and the lmerTest package (Kuznetsova, Bruun Brockhoff, & Haubo Bojesen Christensen, 2016). Across all regressions, we used the theoretical characteristics of the respective distributions as independent variables. We dummy-coded variance and skewness and treated the mean as a continuous predictor variable. As dependent variables, we defined the logarithm of sample size and participant accuracy, quantified as the deviation of their answers proportional to the true means of the distributions. The reason we chose this accuracy measure was to prevent heteroscedasticity in the data, that is, to prevent high answers from having a stronger influence on the regression results than low answers.¹

2.2 | Results

On average, participants drew $M = 28.27$ samples from each distribution ($Mdn = 21$, $SD = 25.62$). There was no significant difference in sample size between estimation ($M = 30.38$, $Mdn = 23$, $SD = 29.95$) and valuation ($M = 26.15$, $Mdn = 22$, $SD = 20.11$). Participants in

¹ Regression analyses based on actual samples led to qualitatively similar results and are reported in the Supporting Information.

FIGURE 1 Schematic for one trial in the estimation (left) and valuation (right) tasks in Experiments 1A, 1B, and 2. Participants sampled from the white box and typed their answer into the gray box. In this example trial, the participant sampled 127



both tasks, however, sampled more when the variance was higher ($b = 0.14, p < .001$).² This is in line with previous findings in the literature (Ashby, 2017; Lindskog et al. 2013), and it is adaptive in the sense that more samples mitigate higher uncertainty. Finally, the higher the mean, the fewer the samples taken ($b = -0.001, p < .01$). This effect can be explained by the decrease in relative variation with higher means and is discussed below for valuations and estimations.

2.2.1 | Valuation task

Figure 2 (left) plots the proportional deviation of participant answers from the true means across the different distributional characteristics in the valuation task. Participants gave certainty equivalents lower than the true means, and the average deviation from the true mean was $M = -4.77$ ($Mdn = -3.13, SD = 16.28$). This was corroborated by a t test on the level of participant means, $t(50) = -3.10, p = .003$. To test the robustness of this inference, we also calculated a Wilcoxon test that led to the same conclusion ($W(n = 51) = 339, p < .001$).

Higher variance sequences led to lower certainty equivalents ($M_{high} = -7.04, Mdn_{high} = -6.25, SD_{high} = 21.49$) than those of lower variance ($M_{low} = -2.44, Mdn_{low} = -2.50, SD_{low} = 7.31$). The left column of Table 2 shows the corresponding regression results for the valuation task. In particular, the parameter for variance is negative ($b = -4.43, SE = 0.69$), that is, higher variance led to significantly lower valuations. Together with the result of overall undervaluation of the mean, these results are consistent with risk-averse preferences.

Skewness also had a significant effect on economic valuations. Participants gave lower values to left- ($M_{left} = -7.26, Mdn_{left} = -3.13, SD_{left} = 17.66$) than to right-skewed distributions ($M_{right} = -2.21, Mdn_{right} = -3.13, SD_{right} = 14.62$). In line with this, the regression analysis has shown that left-skewed outcome distributions were val-

ued significantly lower than normally distributed ones ($b = -2.57, SE = 0.84$) and that right-skewed outcome distributions were valued significantly higher than normally distributed ones ($b = 2.60, SE = 0.84$). Thus, participants preferred right-skewed distributed outcomes.

Finally, deviations got smaller as the mean increased ($b = 0.04, SE = 0.01$). This might be because the variation relative to the mean decreased with higher means when the variance was held constant. We address this issue in Experiment 2.

2.2.2 | Estimation task

The mean estimates within each condition are depicted in Figure 2 (right). As in the valuation task, participants underestimated the theoretical mean of the number sequences across all distributional characteristics ($M = -1.59, Mdn = 0, SD = 9.40$). A t test revealed that this underestimation was significant, $t(50) = -5.17, p < .001$. Again for robustness, we calculated a Wilcoxon test, and it led to the same conclusion ($W(n = 51) = 189, p < .001$).

Underestimation was more pronounced for sequences with high variance ($M_{high} = -2.42, Mdn_{high} = -1.25, SD_{high} = 12.95$) than for those with low variance ($M_{low} = -0.76, Mdn_{low} = 0, SD_{low} = 2.81$). Variance was a significant predictor for estimation deviations in the regression ($b = -1.66, SE = 0.59$; Table 2 right-middle column).

Furthermore, mean estimates for left-skewed distributions ($M_{left} = -3.41, Mdn_{left} = -1, SD_{left} = 8.70$) were lower than for right-skewed distributions ($M_{right} = 0.37, Mdn_{right} = 0, SD_{right} = 7.52$). The regression results revealed that mean estimates of right-skewed distributions were significantly higher ($b = 2.11, SE = 0.64$), and mean estimates of left-skewed distributions significantly lower ($b = -1.68, SE = 0.64$), than mean estimates of normally distributed sequences. Together, all these effects are in accordance with a compressed mental number line.

Finally, the proportional deviation from the theoretical mean got smaller with higher means ($b = 0.02, SE = 0.01$). As in the valuation

²Results are based on a mixed-effects regression with log sample size as the dependent variable. Detailed results can be found in the Supporting Information.

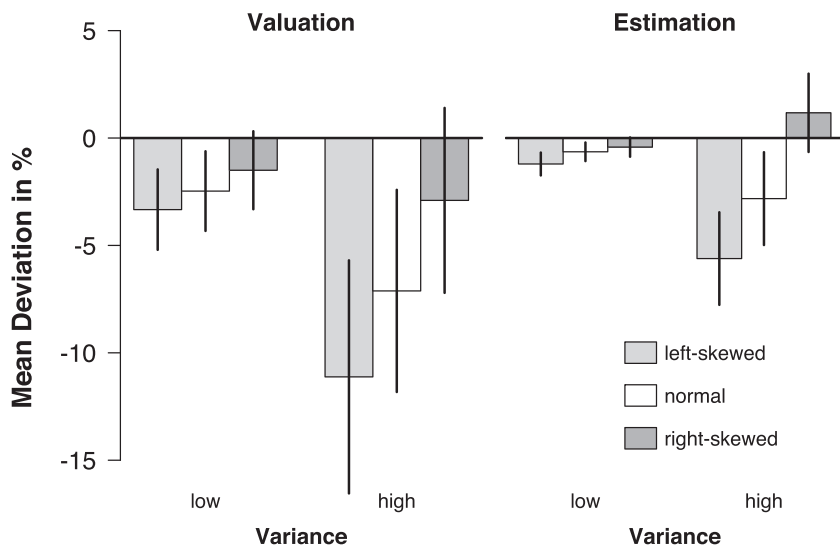


FIGURE 2 Answers in Experiment 1A. The y axis shows percentage deviation of participant answers from the theoretical means of the distributions. Error bars are 95% confidence intervals

	Valuation 1A	Valuation 1B	Estimation 1A	Estimation 1B
(Intercept)	-7.422** (2.119)	-5.594** (2.070)	-3.780** (1.152)	-2.090 (1.291)
Mean	0.041*** (0.011)	0.024* (0.011)	0.024** (0.009)	0.004 (0.007)
SD 20	-4.432** (0.685)	-4.897** (0.687)	-1.660** (0.523)	-0.331 (0.453)
Right skewed	2.598** (0.838)	3.422*** (0.842)	2.108*** (0.640)	1.400* (0.553)
Left skewed	-2.573** (0.838)	-2.213** (0.842)	-1.675** (0.640)	-0.755 (0.553)

TABLE 2 Experiments 1A and 1B: answers in valuation and estimation tasks

Note: Effects of theoretical mean, variance, and skewness on percentage deviation of answers from the theoretical mean in economic valuation and estimation for Experiments 1A and 1B. All models included subject random intercepts. All significant predictors are robust to the inclusion of random slopes. Standard errors in parentheses. * $p < .05$. ** $p < .01$. *** $p < .001$.

task, this effect was most likely due to a decrease in relative variation as the mean increased.

3 | EXPERIMENT 1B

Experiment 1B was a direct, preregistered replication of Experiment 1A (<https://osf.io/ex3dy/>).

3.1 | Method

The main difference was in participant instructions, as participants in Experiment 1A indicated some difficulty in comprehending the incentive scheme (particularly the BDM auction). Hence, in Experiment 1B, we simply instructed participants to answer thoroughly and stated that their accuracy would influence their final payoff. We further informed participants that details of the actual payment mechanism were available upon clicking an extra button on the screen. About one third of participants in each block made use of this option.

3.1.1 | Participants

We tested 58 participants from the same subject pool as in Experiment 1A. Participants received course credit and a choice-dependent bonus

of 1.50 to 8.93 AUD ($M_{\text{pay}} = 5.43$ AUD). Prior to analyzing the data, we excluded answers more than five standard deviations from the distribution's mean (33 out of 2784 total trials) as preregistered. We used data analyses as preregistered, and deviations were clearly marked.

3.2 | Results

On average, participants drew $M = 29.29$ samples from each distribution ($Mdn = 19, SD = 35.49$). More samples were drawn in valuation ($M = 31.71, Mdn = 20, SD = 41.71$) than in estimation ($M = 26.83, Mdn = 19, SD = 27.61$).³ Unlike in Experiment 1A, the mean had no significant influence on the sample size. However, in line with Experiment 1A, participants sampled more when variance was higher ($b = 0.18, p < .001$).

3.2.1 | Valuation task

Figure 3 (left) plots the proportional deviation of participant answers from the true means in the valuation task. All effects were qualitatively

³This effect was significant in a regular regression on individually standardized sample sizes, but failed to reach significance in a mixed-effect regression. For details, see the Supporting Information.

similar to Experiment 1A. Overall, valuations were below the mean ($M = -4.80$, $Mdn = -3.13$, $SD = 17.15$), lower for high- than for low-variance sequences ($M_{high} = -7.25$, $Mdn_{high} = -6.25$, $SD_{high} = 23.11$; $M_{low} = -2.34$, $Mdn_{low} = -1.88$, $SD_{low} = 6.42$), and lower for left- than for right-skewed sequences ($M_{left} = -7.40$, $Mdn_{left} = -2.31$, $SD_{left} = 19.86$; $M_{right} = -1.76$, $Mdn_{right} = -3$, $SD_{right} = 12.44$). A t test and mixed-effect regressions show that certainty equivalents were significantly lower than the theoretical means, $t(57) = -3.26$, $p = .002$,⁴ and higher variance led to significantly lower valuations ($b = -4.90$, $SE = 0.69$). In addition, left-skewed outcome distributions were valued lower than normally distributed ones ($b = -2.21$, $SE = 0.84$), and right-skewed outcome distributions were valued higher than normally distributed ones ($b = 3.42$, $SE = 0.84$). Finally, the mean again had a significantly positive effect ($b = 0.02$, $SE = 0.01$) on participant valuations.

3.2.2 | Estimation task

Figure 3 (right) plots the proportional deviation of participant answers from the true means in the estimation task. All effects trended in the same directions as in Experiment 1A and as in the valuation task. In particular, overall mean estimations were below the mean ($M = -1.22$, $Mdn = 0$, $SD = 9.72$), lower for high- than low-variance sequences ($M_{high} = -2.07$, $Mdn_{high} = -0.77$, $SD_{high} = 13.12$; $M_{low} = -0.70$, $Mdn_{low} = 0$, $SD_{low} = 3.01$), and lower for left- than for right-skewed sequences ($M_{left} = -2.17$, $Mdn_{left} = 0$, $SD_{left} = 9.03$; $M_{right} = -0.04$, $Mdn_{right} = 0$, $SD_{right} = 10.85$). However, the mean deviation in estimation was only marginally significantly lower than zero, $t(57) = -1.80$, $p = .078$,⁵ and higher variance had no significant effect ($b = -0.33$, $SE = 0.45$). Whereas left-skewed distributions were not estimated significantly lower than normal ones ($b = -0.75$, $SE = 0.55$), right-skewed distributions were estimated significantly higher than normal ones ($b = 1.40$, $SE = 0.55$). Finally, the mean had no significant effect on participant estimates ($b = 0.004$, $SE = 0.007$).

3.3 | Discussion of Experiments 1A and 1B

Experiment 1A and its replication in Experiment 1B showed robust effects of overall undervaluation as well as lower valuations for high versus low variance and lower valuations for left- versus right-skewed distributions. These findings refute the idea that rare events are underweighted in the context of continuous outcome distributions, in that, if people underweighted rare events, they would value left- more than right-skewed distributions. The evidence for overweighting high numbers is mixed. The overweighting of high numbers is consistent with higher valuations for right- than for left-skewed distributions, but it is not in line with overall risk-averse valuations. Finally, all these effects align with the predictions of a compressed mental number line.

The qualitative effects in the estimation task were similar to those in the valuation task. In particular, the positive effect of skewness on estimations was robust in Experiments 1A and 1B. Presumably due to

a smaller effect size, underestimation of the mean and the influence of variance on estimations was not always significant.

Surprisingly, increasing the mean led to less undervaluation and trended toward less underestimation, which was not in line with any of the theories reviewed above. A possible confound could be that we kept the variances of the underlying distributions constant across all mean levels. Thus, there was proportionally less variation for high- than for low-mean sequences (see also Whalen et al. 1999; Weber et al. 2004). To further clarify these issues, we conducted another experiment, outlined next.

4 | EXPERIMENT 2

Experiment 2 aims to clarify the influence of the absolute variance versus the coefficient of variation (relative to the EV) on valuations and estimations (Weber et al. 2004). Therefore, in contrast to the previous experiments, we held the coefficient of variation constant over a wide range of EVs. We also increased the number of participants to clarify the ambiguous effects of overall underestimation and of lower estimates for high variance.

4.1 | Method

4.1.1 | Material

We made several changes from the previous experiments. We constructed distributions by holding the coefficient of variation (standard deviation/mean) constant across different mean levels. We introduced three variation levels (5%, 10%, and 20%) and eight mean levels (30, 50, 75, 100, 130, 160, 200, and 250). To keep from inflating the number of trials, we omitted the skewness manipulation. Finally, we drew the offer for the BDM auction in the valuation task from 0% to the 99% quantile of the respective distribution.

4.1.2 | Participants and procedure

We recruited 120 participants from the University of Geneva subject pool. We determined the sample size before the start of the experiment and doubled it from the previous experiments to increase power. The experiment was conducted on single computer workstations and lasted 30 min on average. The participants' average age was 23 years ($Mdn = 22$, $SD = 5.65$), with 40 males and 80 females. In addition to the show-up fee of 20 CHF, the average choice-dependent bonus was approximately 6 CHF, $range = [1.05; 18.25]$.

4.2 | Results

We excluded trials in which responses were more than five standard deviations from the true mean. In contrast to the previous experiments, we also excluded six participants with more than five such trials in either condition, as such behavior indicates a general misunderstanding. However, all reported effects were robust even when the trials of these participants that were within five standard deviations were included.

On average, participants drew $M = 28.68$ samples from each distribution ($Mdn = 20$, $SD = 29.40$). There was no difference in sample size between task types (estimation: $M = 29.86$, $Mdn = 20$, $SD = 29.68$,

⁴For robustness, we also report the result of a Wilcoxon test: $W(n = 58) = 465$, $p = .003$.

⁵For robustness, we also report the result of a Wilcoxon test that was not preregistered: $W(n = 58) = 541$, $p = .015$.

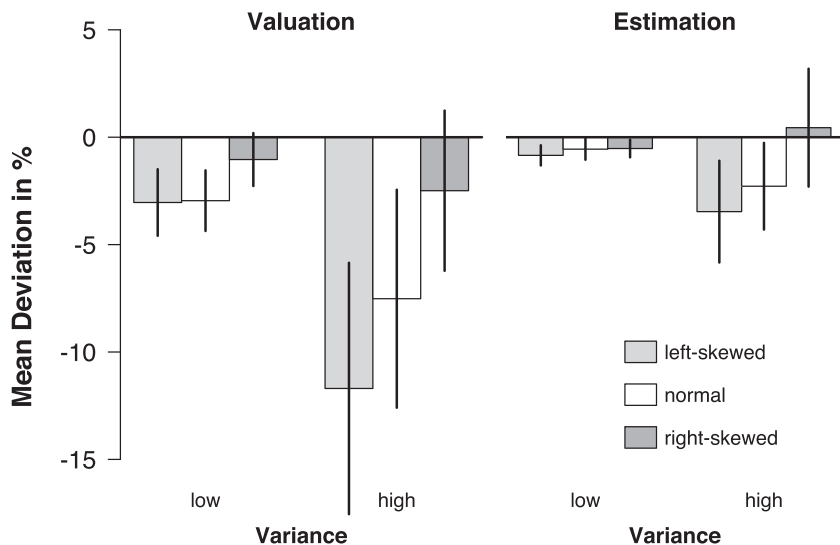


FIGURE 3 Answers in Experiment 1B. The y axis shows percentage deviation of participant answers from the theoretical means of the distributions. Error bars are 95% confidence intervals

valuation: $M = 27.50$, $Mdn = 20$, $SD = 29.06$). Similar to the previous experiments, sample size increased with the variation (10%: $b = 0.09$, $SE = 0.01$, 20%: $b = 0.16$, $SE = 0.02$, each compared to 5% variation). In contrast to the previous experiments, sample size also increased with higher means ($b = 0.0007$, $SE = 0.0001$). We attribute the latter effect to the change in design to hold the coefficient of variation, rather than the variance, constant across mean levels.

4.2.1 | Valuation task

Figure 4 (left) plots the proportional deviation of participant answers from the true means in the valuation task. Participants gave certainty equivalents lower than the true means. The average deviation from the theoretical mean was $M = -5.34$ ($Mdn = -2.80$, $SD = 13.63$). A t test ($t(113) = -6.05$, $p < .001$) and a Wilcoxon test ($W(n = 114) p < .001$) showed that certainty equivalents were significantly lower than the theoretical means.

Higher variation relative to the mean led to lower certainty equivalents than for sequences with lower variation (5%: $M = -1.71$, $Mdn = 0$, $SD = 4.92$; 10%: $M = -3.06$, $Mdn = -1.33$, $SD = 9.39$; 20%: $M = -5.22$, $Mdn = -2.67$, $SD = 16.34$). Regression results (middle column of Table 3) showed that variation was a significant negative predictor for the two higher levels of variation (10%: $b = -1.96$, $SE = 0.46$, 20%: -5.51 , $SE = 0.46$).

Because we explicitly manipulated variation, we can now differentiate the effect of variation from the effect of higher means. There was no significant effect of the mean level ($b = 0.001$, $SE = 0.002$) on participant valuations. This indicates that the effect of the mean on valuation that we found previously was, indeed, due to the confound with the variance.

4.2.2 | Estimation task

The proportional deviations of the mean estimates are depicted in Figure 4 (right). With increased power, we replicated the finding of Experiment 1A that participants underestimated the theoretical mean of the number sequences across all distributions ($M = -1.34$, $Mdn = 0$, $SD = 8.03$). Both a t test ($t(113) = -4.42$, $p < .001$) and a Wilcoxon test ($W(n = 114) = 1383$, $p < .001$) revealed significant underestimation.

TABLE 3 Experiment 2: answers in valuation and estimation tasks

Distribution characteristic	Valuation	Estimation
(Intercept)	-2.99** (0.98)	-0.78 (0.44)
Mean	0.001 (0.003)	0.0001 (0.002)
Variation 10%	-1.96*** (0.46)	-0.68 (0.35)
Variation 20%	-5.51*** (0.46)	-1.40*** (0.35)

Note: Effects of theoretical mean and variance on percentage deviation of answers from the theoretical mean in economic valuation and estimation. All models include subject random intercepts. All significant predictors are robust to the inclusion of random slopes. Standard errors in parentheses. * $p < .05$. ** $p < .01$. *** $p < .001$.

Similarly, we found that higher variation relative to the mean led to lower estimates (5%: $M = -0.65$, $Mdn = 0$, $SD = 3.55$; 10%: $M = -1.31$, $Mdn = 0$, $SD = 7.06$; 20%: $M = -2.06$, $Mdn = 0$, $SD = 11.37$). Regression results (right column of Table 3) show that variation was a negative predictor for the highest level of variation (10%: $b = -0.68$, $SE = 0.35$, 20%: $b = -1.40$, $SE = 0.35$).

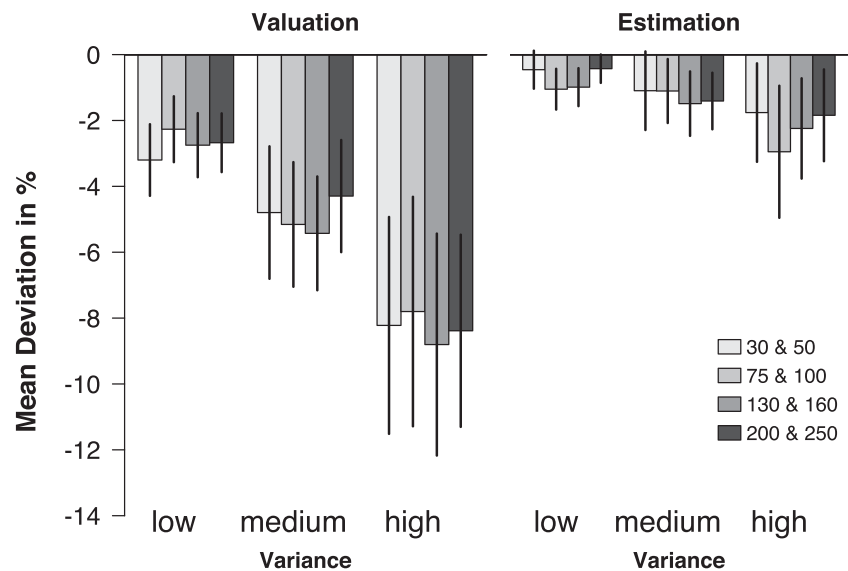
As in the valuation task, the proportional deviation from the theoretical mean was not affected by higher means ($b = 0.0001$, $SE = 0.002$). Again, this indicates that the effect of the mean on estimation in Experiment 1A was spurious.

4.3 | Discussion of Experiment 2

We replicated all effects in valuations that aligned with the predictions of a compressed mental number line with a higher number of participants. Furthermore, when we held the coefficient of variation constant, the mean did not influence valuation or estimation. This shows that the coefficient of variation, and not absolute variance alone, affected valuations (Weber et al. 2004).

Moreover, we replicated the main effect of overall underestimation and the effect of lower estimations for distributions with higher variation, as reported in Experiment 1A, with a higher number of par-

FIGURE 4 Answers in Experiment 2. The y axis shows percentage deviation of participant answers from the theoretical means of the distributions. The legend refers to the means of the distributions. Error bars are 95% confidence intervals



participants. Together with the skewness effect replicated in Experiment 1B, the observed pattern in the estimation task is robust and qualitatively similar to the pattern of the valuation task. This shows that the observed behavior extends qualitatively from preferential to nonpreferential tasks. The compressed mental number line hypothesis—which is not based on preferences—is consistent with this observation.

5 | EXPERIMENTS 3A AND 3B

In the previous studies, the number of samples from each distribution was determined by the participants themselves. Although this aligns with the typical experimental design in the DFE literature, possible behavioral effects in estimation and valuation could be influenced by the number of samples or endogenous stopping. For example, if participants stop sampling after seeing an extremely low number in a left-skewed distribution, this could (partly) cause the effect of skewness on estimation and valuation. To control for these effects, we employed a fixed sampling design where every participant saw the same number of samples for every number sequence.

Experiment 3A aimed to better understand if and how numeric cognition affects economic valuation. Therefore, we included one condition where we told participants the true mean of the sampled distribution prior to their valuation and another condition where participants first made an estimation and then a valuation directly afterwards. If number perception and integration directly affect valuation, then knowing the (unbiased) mean should lead to less negative deviations from the mean than in the condition where participants have only their (biased) mean estimate available. Likewise, if number perception and integration affect estimation and valuation similarly, then valuation answers should correlate with answers in the previous estimation task.

Experiment 3B further examined the hypothesis that individuals have a tendency to simultaneously underestimate and undervalue the observed outcome distributions. Additionally, we assessed the influence of individual numeracy ability on underestimation and undervaluation. In Experiment 3B, estimations and valuations were both based on the same sequences, but unlike Experiment 3A, the

sequences were presented in different blocks. Both experiments were preregistered (<https://osf.io/76t9y> and <https://osf.io/tgrm6>). All details of the methods and the complete results can be found in the Supporting Information.

5.1 | Method

In both Experiments 3A and 3B, participants drew the fixed number of 20 samples from each distribution. Participants were recruited online via Prolific ($n = 131$ and $n = 133$). For the payment of the valuation task, the BDM offer was randomly drawn from the lowest outcome to the highest outcome of the selected sequence.

Sequences in Experiment 3A varied in mean (80–120), standard deviation (10 vs. 20), and skewness (−2.5 vs. 0 vs. +2.5). There were two experimental conditions: in one, participants estimated the mean before making a valuation; in the other, participants were told the true mean before making a valuation.

In Experiment 3B, only normally distributed sequences were presented. These distributions had similar variance but different means. At the end, participants completed the Berlin Numeracy Test (Cokely et al. 2012), which consisted of four questions about calculation with probability.

5.2 | Results and discussion

Using a fixed sampling design, Experiment 3A replicated the effects of underestimation and undervaluation, as well as the negative effect of variance and the positive effect of skewness, in both tasks. This showed that previous results did not depend on motivated sampling or endogenous sample size. Undervaluation was less pronounced when the true mean was known than when it was unknown ($M = -0.37\%$, $Mdn = -0.28\%$, $SD = 11.34\%$; regression: $b = -0.21$, $p < .001$). Knowing the mean led to less undervaluation by about 3.39% of that in the condition where the mean was not known. In addition, the deviation in the estimation condition predicted the deviation in the following valuation, $b = 0.62$, $p < .001$. As a limitation, these effects might also have been driven by the fact that valuation followed directly

after true-mean presentation or estimation, which may have set an anchor for the valuation.

In Experiment 3B, we again replicated the effects of underestimation and undervaluation. When estimation and valuation tasks were elicited in different blocks and based on separate (but identical) samples, there was no significant correlation between the mean of percentage deviations for each participant in both tasks ($r = .11$, $p = .13$). Further, the Spearman correlation between the numeracy score and the participant mean deviation was $r = .13$ ($p = .076$) in the estimation task and $r = -.05$ ($p = .710$) in the valuation task. We conclude that in this experiment the evidence for an individual tendency to underestimate and undervalue remains inconclusive. Overall, trial-by-trial variability seemed to be very strong, which impeded the correlation of individual differences between tasks (Rouder, Kumar, & Haaf, 2019).

6 | SUMMARY OF RESULTS

To summarize the results across all experiments, we combined and analyzed the data in an internal meta-analysis. The overall sample size was 454, and all previously reported effects were significant for both valuation and estimation. Estimation effects were smaller than valuation effects but were always in the same direction. To quantify the degree to which estimation resembled valuation, we used the ratio (estimation/valuation) of each respective effect.

Across all experiments, the overall ratio of underestimation to undervaluation was 19%. Taking the difference in answers between low- ($SD = 5$) and high-variance ($SD = 20$) sequences separately for the two tasks yielded a ratio of 27%. Finally, taking the difference in answers between left- and right-skewed distributed sequences separately for the two tasks yielded a ratio of 48%. This means that the greatest similarity in the size of the effect between the estimation and valuation tasks was found in the effect of the skewness.⁶ The ratios across all experiments, as well as separately for each experiment where the particular effect was manipulated, are presented in Figure 5.

7 | GENERAL DISCUSSION

In examining experience-based valuations of gambles with continuous outcome distributions, we found that people showed risk aversion. They gave valuations below the sequence means, and valuations were lower for high- than for low-variance sequences. In addition, participants valued right-skewed higher than left-skewed distributions. These results disconfirm the hypotheses that underweighting rare events and overweighting extreme events generalize to valuations of continuous outcomes. We found a similar qualitative, though less pronounced, pattern in a task where participants estimated the mean of a sequence, and economic preferences for risk or skewness should play no role. This shows that the characteristics of a number distribution affect the perception and integration of numeric information beyond the narrow area of economic valuation under risk and uncertainty. The behavioral effects, as well as the similarity between valuations

and estimations, are consistent with the idea that numeric cognition is subject to a compressed mental number line (Feigenson et al. 2004). Consequently, economic valuations of gambles could be shaped partly (20% to nearly 50%, according to our data) by regularities in numeric cognition rather than by subjective preferences alone.

7.1 | Valuations versus choices from experience

The main difference between our study and those in the DFE literature is that we elicited valuations of single number sequences, whereas most previous studies asked people to choose between two number sequences. In choice studies, people have tended to prefer higher variance sequences, a behavior explained by the overweighting of high numbers (Glickman et al. 2018; Konstantinidis et al. 2018; Ludvig & Spetch, 2011; Spitzer et al. 2017; Tsetsos et al. 2012). This contrasts with the lower valuations of high-variance sequences observed in our experiments. Future research is needed to shed more light on this gap between valuations and choice in DFE (see Golan & Ert, 2015). Crucial differences between the two paradigms that could be relevant are the number of streams the decider has to pay attention to (Vanunu, Pachur, & Usher, 2019), the goal of the decider to come up with either a precise monetary amount or to make an ordinal comparison, and the attitude of the decider toward perceived (relative) losses (Ashby et al. 2018; Kunar et al. 2017).

Another regularity often found in choice studies of DFE is that people behave as if they underweight rare events (Hertwig et al. 2004). This contrasts with the higher valuations observed in our experiments for right-skewed distributions, with rare high outcomes, than for left-skewed distributions, with rare small outcomes. Prior choice studies that report underweighting of rare events have usually used situations with one safe option and another option with two outcomes. When participants have looked at pairwise options between two-outcome gambles, their choices were more in line with attenuated overweighting of rare events (Glöckner et al. 2016; Kellen et al. 2016). In addition to the difference in elicitation format, unlike previous studies, our experiments used continuous outcome sequences and, thus, a different definition of rare events.

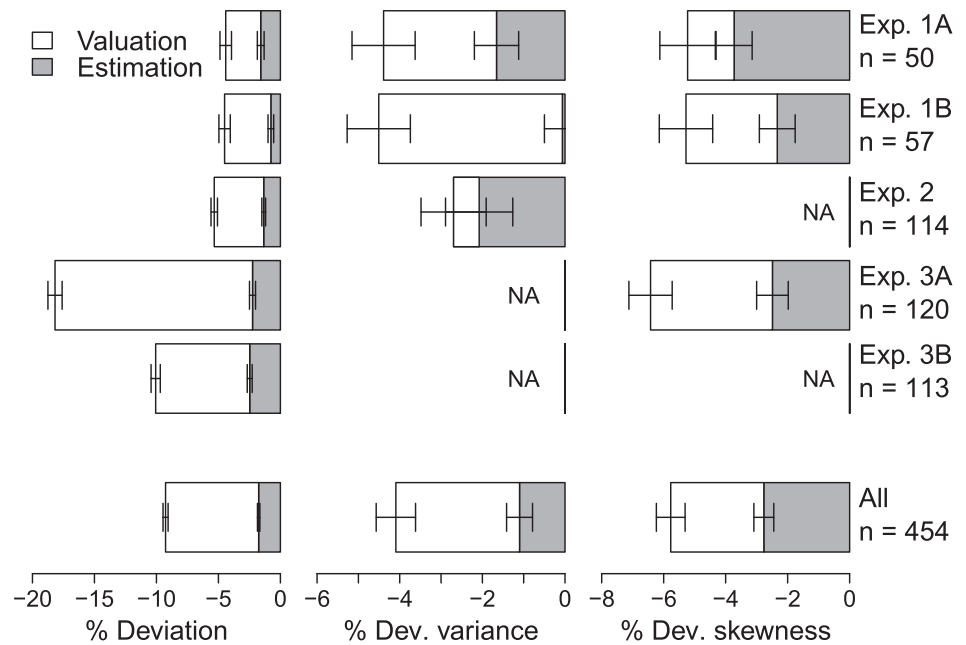
7.2 | Valuations from experience and numeric cognition

We found a statistically robust but small underestimation of the mean of -1.72% across all experiments relative to the actual means of the observed number sequences. This is consistent with recent research (Brezis et al. 2015; Scheibehenne, 2019). Yet older studies did not find such an effect and described people as intuitive statisticians (e.g., Beach & Swenson, 1966; Laestadius Jr, 1970; Spencer, 1963). A possible explanation for this discrepancy is that older studies focused primarily on absolute or squared deviations from the mean and that way did not measure estimation biases. In addition, older studies typically had smaller numbers of participants; thus, the small effect size of the underestimation might not have reached statistical significance.

Biased estimations of number sequences could result from an intuitive number sense that guides the perception and integration of numbers (Brezis et al. 2015; Feigenson et al. 2004; Gallistel & Gelman, 2000). Because number perception and integration are part

⁶We do not report the ratio for the mean, because as shown in Experiment 2, this effect vanishes once we hold the variation (instead of the variance) constant.

FIGURE 5 Relation of effects in estimation and valuation across all experiments. The x axis shows the percentage deviation from the mean, the difference in percentage deviation between trials with $SD = 20$ and $SD = 5$, and the difference in percentage deviation between trials with left- and right-skewed distributions. In all experiments, participants with more than five estimates further away than ± 5 SD from the true mean were removed. Results are shown only for experiments where the respective manipulation was present (otherwise, NA). The column “All” shows the average values from all experiments where data are available. Error bars are standard errors of the mean



of economic valuation, this could explain similar qualitative patterns in estimation and valuation (Schley & Peters, 2014). The mental number line provides the best explanation for the effects we have observed, as compared with other accounts derived from the DFE literature. Whereas past research on the mental number line has predominantly focused on nonsymbolic numerosities such as dot clouds (e.g. Dehaene et al. 2008), our study is consistent with compression in symbolic numbers. Future studies are needed to examine the differences and similarities in compression between streams of symbolic numbers and nonsymbolic numerosities in DFE (see Dutilh & Rieskamp, 2016; Zeigenfuss et al. 2014).

As a limitation, we cannot conclude from the similarity between estimation and valuation that mean estimates are direct antecedents of valuations nor that valuations are causally influenced by numeric cognition. For one, the individual correlation between estimation and valuation was surprisingly small given the similar patterns in the aggregate. Further, there are alternative explanations for the data. For example, the effects in Experiment 3A could be due to anchor effects or due to common-method biases, a task characteristic that has affected both estimation and valuation similarly (Podsakoff, MacKenzie, & Podsakoff, 2012). Moreover, we are not claiming that estimation and valuation are identical. They are clearly different in that all effects in estimation were smaller in size than the effects in valuation. This difference could stem from the increased importance of distributional characteristics like variance or skewness in preferential tasks. To rigorously establish a causal link between numeric cognition and economic valuation, a direct manipulation of the underlying cognitive processes would be required.

7.3 | The cognitive underpinning of preferences

The current research aims to add to the literature that explores the cognitive underpinnings of economic preferences (Dertwinkel-Kalt & Köster, 2020; Khaw et al. 2017; Lieder et al. 2018; Schley & Peters, 2014). The DFE paradigm seems particularly useful to a better

understanding of how numeric information is processed to come up with preferential decision making. One defining characteristic of this task is that numeric information is presented sequentially and must be stored in memory as it is usually not available at the time a decision is made. Hence, memory effects can mediate behavioral characteristics as, for example, the overweighting of extreme events (Kahneman et al. 1993; Madan et al. 2014). The process of sequential number integration can also be explicitly modeled through online-updating or memory-based individual number recall (Erev et al. 2008; Mason et al. 2019; Gonzalez et al. 2003). Future research is needed to link these memory models to the compressed mental number line and to examine whether compression differs depending on the number of memory processes that are necessary to integrate numeric information. Another class of cognitive models, range-frequency theory and decision-by-sampling, examines how the distribution of individual numbers shapes the perception of individual and summary evaluations (Parducci et al. 1968; Stewart et al. 2006; Tripp & Brown, 2016). In sum, a better understanding of how the context and memory processes influence the valuation of numbers can inform models of preferential decision making.

In general, processing numeric information requires cognitive resources. To the extent that people differ in their cognitive abilities, this could also affect preferential decision making (Ashby, 2017; Dohmen et al. 2018). In the current studies, we found no conclusive evidence of a relation between underestimation and numeracy. This could be due to the small effect size of this correlation. However, it could also mean that more refined cognitive models are necessary to find a direct relation between cognitive ability and DFE. One important point is that the numeracy questionnaire mostly asks about the understanding of probabilities, whereas no explicit probabilities have to be calculated in DFE. Further, we are not aware of any research linking the curvature of the mental number line to numeracy explicitly. Other measures of cognitive ability may be more important to understanding individual differences in DFE. Given the memory component present in DFE, a plausible candidate could be working memory capacity, that

is, the ability to store and manipulate items in short-term memory (Frey, Mata, & Hertwig, 2015).

Finally, many important laboratory studies about economic behavior make use of the DFD paradigm (e.g., Holt & Laury, 2002; Tversky & Kahneman, 1992). Cognitive models could also help us better understand behavioral differences between DFD and DFE by examining the process of numeric integration. Whereas outcomes must be integrated with probabilities when information is summarized descriptively, single outcomes must be integrated when information is presented sequentially. Thus, differences in the context of numbers presented and the involvement of memory processes in DFE could lead to divergent behavior in both paradigms. Future research is needed to synthesize the above points through the integration of cognitive models into the examination of preferential behavior in both DFD and DFE.

ACKNOWLEDGEMENTS

This research was supported by Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (SNF) grant P2BSP1_188188 to the first author and SNF grants 100014_149846 and 100014_175849 to the last author.

DATA AVAILABILITY STATEMENT

Supplemental material and the data are available at the Open Science Framework: <https://osf.io/ehkuz/>.

ORCID

Sebastian Olschewski  <https://orcid.org/0000-0001-9371-1597>

REFERENCES

- Ashby, N. J. S. (2017). Numeracy predicts preference consistency: Deliberative search heuristics increase choice consistency for choices from description and experience. *Judgment and Decision Making*, 12(2), 128.
- Ashby, N. J. S., & Rakow, T. (2014). Forgetting the past: Individual differences in recency in subjective valuations from experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(4), 1153.
- Ashby, N. J. S., Yechiam, E., & Ben-Eliezer, D. (2018). The consistency of visual attention to losses and loss sensitivity across valuation and choice. *Journal of Experimental Psychology: General*, 147(12), 1791.
- 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.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Baucells, M., Weber, M., & Welfens, F. (2011). Reference-point formation and updating. *Management Science*, 57(3), 506–519.
- Beach, L. R., & Swenson, R. G. (1966). Intuitive estimation of means. *Psychonomic Science*, 5(4), 161–162.
- Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, 9(3), 226–232.
- Bordalo, P., Gennaioli, N., & Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly Journal of Economics*, 127(3), 1243–1285.
- Brezis, N., Bronfman, Z. Z., & Usher, M. (2015). Adaptive spontaneous transitions between two mechanisms of numerical averaging, 5, 10415. *Scientific Reports*.
- Cheyette, S. J., & Piantadosi, S. T. (2019). A primarily serial, foveal accumulator underlies approximate numerical estimation. *Proceedings of the National Academy of Sciences*, 116(36), 17729–17734.
- Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin numeracy test. *Judgment and Decision Making*.
- Dehaene, S. (2003). The neural basis of the Weber–Fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences*, 7(4), 145–147.
- Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigenous cultures. *Science*, 320(5880), 1217–1220.
- Dertwinkel-Kalt, M., & Köster, M. (2020). Salience and skewness preferences. *Journal of the European Economic Association*, 18(5), 2057–2107.
- Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2018). On the relationship between cognitive ability and risk preference. *Journal of Economic Perspectives*, 32(2), 115–34.
- Dutilh, G., & Rieskamp, J. (2016). Comparing perceptual and preferential decision making. *Psychonomic Bulletin & Review*, 23(3), 723–737.
- Erev, I., Glözman, I., & Hertwig, R. (2008). What impacts the impact of rare events. *Journal of Risk and Uncertainty*, 36(2), 153–177.
- Fechner, G. T. (1860). *Elemente der psychophysik*: Breitkopf and Härtel.
- Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, 8(7), 307–314.
- Fox, C. R., & Hadar, L. (2006). “Decisions from experience” = Sampling error + prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004). *Judgment and Decision Making*, 1(2), 159.
- Frey, R., Mata, R., & Hertwig, R. (2015). The role of cognitive abilities in decisions from experience: Age differences emerge as a function of choice set size. *Cognition*, 142, 60–80.
- Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences*, 4(2), 59–65.
- Genest, W., Stauffer, W. R., & Schultz, W. (2016). Utility functions predict variance and skewness risk preferences in monkeys. *Proceedings of the National Academy of Sciences*, 113(30), 8402–8407.
- Glickman, M., Tsetsos, K., & Usher, M. (2018). Attentional selection mediates framing and risk-bias effects. *Psychological Science*, 29(12), 2010–2019.
- 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.
- Golan, H., & Ert, E. (2015). Pricing decisions from experience: The roles of information-acquisition and response modes. *Cognition*, 136, 9–13.
- Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, 27(4), 591–635.
- 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.
- Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5), 1644–1655.
- Hotaling, J., Jarvstad, A., Donkin, C., & Newell, B. R. (2019). How to change the weight of rare events in decisions from experience. *Psychological Science*.
- Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science*, 4(6), 401–405.
- 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. (2017). Risk aversion as a perceptual bias: National Bureau of Economic Research Tech. Rep.

- Konstantinidis, E., Taylor, R. T., & Newell, B. R. (2018). Magnitude and incentives: Revisiting the overweighting of extreme events in risky decisions from experience. *Psychonomic Bulletin & Review*, 25(5), 1925–1933.
- Kunar, M. A., Watson, D. G., Tsetsos, K., & Chater, N. (2017). The influence of attention on value integration. *Attention, Perception, & Psychophysics*, 79(6), 1615–1627.
- Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2016). lmerTest: Tests in linear mixed effects models. <https://CRAN.R-project.org/package=lmerTest>, R package version 2.0-32.
- Laestadius Jr, J. E. (1970). Tolerance for errors in intuitive mean estimations. *Organizational Behavior and Human Performance*, 5(2), 121–124.
- 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.
- Lindskog, M., & Winman, A. (2014). Are all data created equal?—Exploring some boundary conditions for a lazy intuitive statistician. *PloS One*, 9(5), e97686.
- Lindskog, M., Winman, A., & Juslin, P. (2013). Calculate or wait: Is man an eager or a lazy intuitive statistician? *Journal of Cognitive Psychology*, 25(8), 994–1014.
- 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), e20262.
- Madan, C. R., Ludvig, E. A., & Spetch, M. L. (2014). Remembering the best and worst of times: Memories for extreme outcomes bias risky decisions. *Psychonomic Bulletin & Review*, 21(3), 629–636.
- Madan, C. R., Ludvig, E. A., & Spetch, M. L. (2016). The role of memory in distinguishing risky decisions from experience and description. *The Quarterly Journal of Experimental Psychology*, 70(10), 1–33.
- Mason, A., Brown, G., Ward, G., & Farrell, S. (2019). Memory-based and online strategies in retrospective evaluations. <https://doi.org/10.31234/osf.io/uaqs5>
- Menezes, C., Geiss, C., & Tressler, J. (1980). Increasing downside risk. *The American Economic Review*, 70(5), 921–932.
- Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, 215, 1519–1520.
- Parducci, A., Thaler, H., & Anderson, N. H. (1968). Stimulus averaging and the context for judgment. *Attention, Perception, & Psychophysics*, 3(2), 145–155.
- Peirce, J. W. (2007). Psychopy psychophysics software in python. *Journal of Neuroscience Methods*, 162(1), 8–13.
- Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, 68(1), 29.
- Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, 63, 539–569.
- R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>
- RStudio Team (2015). Rstudio: Integrated development environment for R. RStudio, Inc., Boston, MA, <http://www.rstudio.com/>
- Rouder, J., Kumar, A., & Haaf, J. M. (2019). Why most studies of individual differences with inhibition tasks are bound to fail.
- Scheibehenne, B. (2019). The psychophysics of number integration: Evidence from the lab and from the field. *Decision*, 6(1), 61.
- Schley, D. R., & Peters, E. (2014). Assessing economic value symbolic-number mappings predict risky and riskless valuations. *Psychological Science*, 25(3), 753–761.
- Schoemaker, P. J. H. (1982). The expected utility model: Its variants, purposes, evidence and limitations. *Journal of Economic Literature*, 20, 529–563.
- Spencer, J. (1963). A further study of estimating averages. *Ergonomics*, 6(3), 255–265.
- Spitzer, B., Waschke, L., & Summerfield, C. (2017). Selective overweighting of larger magnitudes during noisy numerical comparison. *Nature Human Behaviour*, 1(8), 0145.
- Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology*, 53(1), 1–26.
- Tripp, J., & Brown, G. D. A. (2016). Being paid relatively well most of the time: Negatively skewed payments are more satisfying. *Memory & Cognition*, 44(6), 966–973.
- Tsetsos, K., Chater, N., & Usher, M. (2012). Saliency driven value integration explains decision biases and preference reversal. *Proceedings of the National Academy of Sciences*, 109(24), 9659–9664.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4), 297–323.
- Vanunu, Y., Pachur, T., & Usher, M. (2019). Constructing preference from sequential samples: The impact of evaluation format on risk attitudes. *Decision*, 6(3), 223–236.
- Weber, E. U., Shafir, S., & Blais, A.-R. (2004). Predicting risk sensitivity in humans and lower animals: Risk as variance or coefficient of variation. *Psychological Review*, 111(2), 430.
- Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, 10(2), 130–137.
- 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.
- Wulff, D. U., & Pachur, T. (2016). Modeling valuations from experience: A comment on Ashby and Rakow (2014).
- Zeigenfuse, M. D., Pleskac, T. J., & Liu, T. (2014). Rapid decisions from experience. *Cognition*, 131(2), 181–194.

AUTHOR BIOGRAPHIES

Sebastian Olschewski is a visiting postdoc at Warwick Business School, University of Warwick. His research examines how perception, attention, and learning affect economic decisions and how people deal with limited cognitive resources.

Ben R. Newell is Professor of Cognitive Psychology and Deputy Head of School in the School of Psychology at UNSW Sydney. He has broad interests across a range of judgment and decision-making phenomena. In recent years he has focused on the cognitive processes underlying multi-attribute judgment as well as described and experience-based risky choice.

Yvonne Oberholzer is a PhD candidate at the Cognition and Consumer Behavior Lab at the Karlsruhe Institute of Technology (KIT). Her research focus lies on Numerical Cognition and Decision Making in general.

Benjamin Scheibehenne is a professor for Cognition and Consumer Behavior at the Karlsruhe Institute of Technology (KIT). He is interested in the psychological processes underlying economic behavior.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Olschewski S, Newell B, Oberholzer Y, Scheibehenne B. Valuation and estimation from experience. *J Behav Dec Making*. 2021;1–13. <https://doi.org/10.1002/bdm.2241>