How Basic Cognition Influences Experience-Based Economic Valuation

Sebastian Olschewski*, Ben R. Newell×, & Benjamin Scheibehenne+

* Center for Economic Psychology, Department of Psychology, University of Basel

× School of Psychology, University of New South Wales

+Geneva School of Economics and Management (GSEM), University of Geneva

Author Note

Sebastian Olschewski, Center for Economic Psychology, Department of Psychology, University of Basel, Missionsstrasse 62a, 4055 Basel, Switzerland, sebastian.olschewski@unibas.ch. Ben R. Newell, School of Psychology, University of New South Wales, Sydney 2052, Australia, ben.newell@unsw.edu.au. Benjamin Scheibehenne, Geneva School of Economics and Management (GSEM), University of Geneva, 40 bd du Pont D'Arve, 1211 Geneva, Switzerland, benjamin.scheibehenne@unige.ch. This research was supported by SNF grants 100014_149846 and 100014_175849 to the third author. Raw data are available on the open science framework: enter link here. Correspondence concerning this article should be addressed to Sebastian Olschewski.

Abstract

The perception and integration of sequential numerical information is a common cognitive task. It is a prerequisite for experience-based economic choices, but it is usually not part of economic decision theory. To better understand the process of symbolic number integration and its influence on economic behavior, we performed three experimental studies that examined mean estimates and economic valuations of continuous number distributions. The results indicate that participants valued random number distributions below their respective arithmetic means and valued distributions as lower when their variance increased, indicating risk aversion. A similar though less pronounced pattern also occurred in the matched mean estimation task where accuracy was incentivized and preferences played no role. These patterns suggest that seemingly risk-averse preferences are partly due to cognitive biases when perceiving and estimating numbers. In addition, participants' apparent economic preference for right-skewed outcome distributions could be attributed mainly to estimation biases. We discuss the extent to which the results can be explained based on a compressed mental number line and different sample weighting models. Finally, a new model that can account for the qualitative data pattern and has stronger overweighting of lower than higher numbers as its core feature is developed. Together, our results indicate that basic cognitive processes in perceiving and integrating number sequences play a key role in understanding experience-based economic behavior.

Keywords: decision from experience, BDM auction, risk preference, continuous outcome distributions, estimation bias

How Basic Cognition Influences Experience-Based Economic Valuation

Introduction

How do people perceive and integrate numerical information that is presented sequentially? A better understanding of the underlying cognitive processes is an important question in its own right and has important implications for economic judgment and decision-making. For example, when one thinks about determinants of real-world investment behavior, what usually comes to mind are economic preferences, such as risk, delay, loss, or uncertainty aversion. Yet, subjective valuations of investment options are also influenced by cognitive number perception and processing (Kahneman, 2003; Khaw, Li, & Woodford, 2017; Krajbich, Armel, & Rangel, 2010; Schley & Peters, 2014; Tsetsos, Chater, & Usher, 2012). So when people think about investing in stocks, they might research the history of returns on the stock market. Integrating and estimating a sequence of discrete monetary returns is a complex cognitive task that requires perception, attention, and working memory. In this paper, we aim to understand how people integrate and evaluate the outcome distribution of a continuous number sequence. In addition, we probe to what extent behavioral phenomena that are usually explained by subjective preferences (e.g. risk aversion) might also depend on regularities in the way people perceive and integrate numerical information.

To distinguish between cognitive number processing on the one hand and economic preferences on the other, we designed a series of experiments in which we gave participants identical numerical information while varying the task: An estimation task asking about the mean as an objective characteristic of an outcome distribution should not involve economic preferences. In contrast, eliciting certainty equivalents/ valuations for an outcome distribution requires both assessing objective aspects such as the mean and incorporating one's own subjective economic preference such as risk-aversion. We made use of this difference between estimation and valuation to disentangle economic preferences from the process of number integration: To the extent that economic valuations are based on (potentially distorted) perception and number integration, behavioral patterns in the estimation task should predict answers in the valuation task.

Economic Preferences

A central concept in economic decision making is risk aversion. It describes two commonly observed behavioral phenomena: First, people typically prefer a sure outcome over a gamble with the same expected value (EV). Second, in the case of two (or more) risky gambles with similar EV, people often prefer the one with lower variance. Risk aversion is often mathematically described in terms of a concave utility function, such that high values are relatively more compressed than small values (Pratt, 1964, Rothschild & Stiglitz, 1971, but see: Weber, Shafir, & Blais, 2004). This also leads to the prediction that gambles with higher stakes lead to more risk-aversion than gambles with lower stakes. Only few studies explicitly examined this effect, but mostly found evidence for an increase in risk-aversion with stakes (Binswanger, 1980; Holt & Laury, 2002).

Whereas risk aversion has been invoked to explain economic preferences with respect to outcome variance, empirical evidence suggests that economic preferences are also affected by higher moments of an outcome distribution. One example is the third moment (i.e. skewness; Åstebro, Mata, & Santos-Pinto, 2015; Kraus & Litzenberger, 1976; Spiliopoulos & Hertwig, 2015). To illustrate, Figure 1 shows distributions that are right-skewed (i.e. high outcomes occur with small probability and most samples are below the mean) and left-skewed (i.e. small outcomes occur with low probability and most samples are above the mean). A preference for right-skewed distributions is one way to explain buying lottery tickets and insurance at the same time (Golec & Tamarkin, 1998; Spiliopoulos & Hertwig, 2015). In line with this reasoning, the mean-variance model (Markowitz, 1952) was extended for skewness preferences with an additional parameter (Kraus & Litzenberger, 1976). Likewise, prospect theory (Kahneman & Tversky, 1979) can incorporate skewness preferences with the shape of the probability weighting function. Often, this function is estimated to give higher weights for rare events, for example, high outcomes in a right-skewed and low outcomes in a left-skewed distribution (but see Hertwig, Barron, Weber, & Erev, 2004).

The ontological status of utility-based models in economics has been debated for over a century. It could be either a parsimonious way to mathematically summarize and describe economic behavior, a so-called as-if model (Friedman, 1953), or it could depict a more fundamental basic regularity of human (and nonhuman) number cognition (Kahneman, 2003). To the extent that economic choices depend on how people perceive and integrate numerical information, research in psychophysics and cognitive science should be integrated into theories about economic behavior.

Experienced Outcomes and Numeric Cognition

The cognitive foundation of economic preferences is supposedly relevant in decisions from experience (DFE; Barron & Erev, 2003; Hertwig et al., 2004; Weber et al., 2004). In a DFE experiment, participants typically sample single outcomes from a numerical distribution before making a choice. This paradigm can have higher external validity and is arguably cognitively more demanding compared to a situation where all possible outcomes and probabilities are presented in a descriptive format. Consequently, the DFE paradigm has been used to test several prominent findings such as the influence of subjective probability weighting (i.e. the under- and overweighting of rare events) (e.g. Barron & Ursino, 2013; Hau, Pleskac, Kiefer, & Hertwig, 2008; Ungemach, Chater, & Stewart, 2009). Yet, these studies were confined to test specific economic preferences and typically focused on choices between a gamble with only one nonzero outcome and a certain outcome. When these analyses were extended to incorporate gambles with two nonzero outcomes, empirical results were mixed (Abdellaoui, L'Haridon, & Paraschiv, 2011; Glöckner, Hilbig, Henninger, & Fiedler, 2016). In the experiments presented below, we further generalize the paradigm by using continuous outcome distributions.

Although there are some theories to explain behavior in an experience-based economic context (e.g. Erev, Glozman, & Hertwig, 2008), so far this literature is not well connected to research about the cognition of numbers. In cognitive psychology, there is evidence that people have an inherently imprecise and nonverbal notion of numbers (Gallistel & Gelman, 2000; Whalen, Gallistel, & Gelman, 1999). To some extent, this may also hold true for symbolic numbers (Brezis, Bronfman, & Usher, 2015; Moyer & Landauer, 1967; Schley & Peters, 2014). Given that economic behavior is stochastic (Hey, 1995; Mosteller & Nogee, 1951), imprecise mental representations could be a source of this stochasticity (see Khaw et al., 2017). In the following, we review models of imprecise mental number sense to explore if they predict a systematic distortion when integrating a sequence of symbolic numbers. Several models of number cognition have been proposed in the literature. Here, we focus on two frameworks of number processing and their respective predictions for the perception and integration of number sequences: The compressed mental number line and the unequal weighting of certain numbers in a sequence.

Compressed Mental Number Line

Research by Dehaene and colleagues (Dehaene, 2011; Feigenson, Dehaene, & Spelke, 2004) indicated that the internal representation of numerals can be described as a compressed mental number line. Historically, concave functions and the resulting compression are a ubiquitous modeling approach in psychophysics and experimental psychology and have been found to describe the perception and discrimination of various entities such as weight, length, and brightness (e.g. Fechner, 1860; Stevens, 1957). A compression with respect to samples in a sequence of numbers implies that people underestimate the mean of a number sequence and show stronger underestimation for high variance compared to low variance sequences. Thus, the compression of numerals could be a source of risk-averse behavior in an economic context (Schoemaker, 1982). Furthermore, it has been shown that for compressed power functions (that is a power coefficient between 0 and 1), sequences' means are higher for right- compared to left-skewed distributions (Genest, Stauffer, & Schultz, 2016; Menezes, Geiss, & Tressler, 1980) . Finally, compression should be greater for higher means compared to lower means.

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, 2017), others have found no such evidence (Lindskog & Winman, 2014; Peterson & Beach, 1967). Also, in contrast to the prediction of a compressed mental number line, a recent study found that an increase in variance of a number sequence led to higher mean estimates (Tsetsos et al., 2012). Consequently, it is an open question if and under what circumstances a compressed mental number line applies when integrating symbolic number sequences.

Sample Weighting Function

A different class of models assumes that the weight each number in a sequence receives to determine an overall magnitude judgment like the sum or the mean could be systematically distorted. In research about experience-based choices, models were developed assuming that people underweight rare events, which were defined as events that occur with at most 20% probability (Hertwig et al., 2004). In the case of continuous and symmetric distributions like the normal distribution, rare events occur equally likely for high and low numbers. Therefore, such a weighting scheme would predict no underestimation of normally distributed outcomes as well as no effect of mean or variance. In contrast, in skewed distributions, rare events are more likely to occur on one side of the distribution. Hence, underweighting of rare events predicts lower mean estimates for right-skewed distributions (where high values are rare) compared to left-skewed distributions (where low values are rare) with symmetric distributions in the middle.

As an alternative to looking at rare events, Ludvig and Spetch, 2011 showed that extreme numbers in a sampling sequence get overweighted. Here, extremeness was defined as the largest (absolute) outcome in the samples of a number sequence. For positive continuous outcome distributions this would mean that a sequence's mean should be overestimated and this effect should be stronger for high variance (compared to low-variance) and right-skewed (compared to left-skewed) outcome distributions. In addition, when extremeness is strictly locally defined within a single number sequence, that is numbers from other sequences in the same experiment or in the real world do not affect the process, differences in the mean of a sequence should not affect the estimation.

Empirical evidence for sample weighting models comes from early studies in decision-from-experience (for a review see Wulff, Mergenthaler-Canseco, & Hertwig, 2018). These studies show for example that people predominantly chose gambles with small probabilities of the worse outcomes over the expected value for certain. Yet, it has to be noted that the underweighting of rare events is not the only explanation for these empirical finding and that later studies looking at choices between two two-outcome gambles came to different conclusions (Glöckner et al., 2016; Kellen, Pachur, & Hertwig, 2016). Concerning the overweighting of extreme events, studies showed that people overestimated the frequency of extreme events in a sequence of numbers (Madan, Ludvig, & Spetch, 2014, 2016). In addition, psychophysics research using sound stimuli found that right-skewed distributions lead to higher mean estimates compared to left-skewed distributions (Parducci, Thaler, & Anderson, 1968) and retrospective ratings of pain have been found to strongly depend on the maximum (i.e. extreme) pain endured (Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993). Finally, as already mentioned above, when people choose between two continuous outcome gambles, people chose on average the higher variance gamble with same expected value, which is in line with the idea that people overweight the highest outcomes (Tsetsos et al., 2012).

To summarize the predicted effects of the respective theories of number cognition on economic behavior: The compressed mental number line predicts overall risk-aversion and a preference for right-skewed over left-skewed gambles as well as an increase in risk-aversion as the stakes (i.e. the mean of a gamble) rise. Sample weighting comes in two flavors: Underweighting of rare events predicts an economic preference for left- over right-skewed distributions but no overall risk-aversion for normally distributed gambles. In contrast, overweighting of extreme events predicts a preference for right-skewed over left-skewed gambles. It further predicts risk-seeking behavior for gains. To systematically test these predictions, we conducted three experiments. Participants in all experiments repeatedly sampled numbers from different continuous payoff distributions and then estimated the mean of the observed number sequence and provided an economic valuation. In contrast to economic valuations, estimations of objective criterion values such as the mean should not be influenced by economic preferences. This paradigm bridges the literature on economic preferences with the literature on cognitive number processing.

Experiments 1 and 2

Method Experiment 1

The Tasks. The experimental tasks were based on the DFE paradigm, where people can freely sample from number distributions for as long as they want before finally making one consequential choice. To assess participants' economic valuations, we had them repeatedly state their certainty equivalents for different outcome distributions from which they could sample. We made the certainty equivalents incentive compatible by asking for minimum selling prices (willingness-to-accept, WTA). It was explained to participants that the minimum selling price is the minimum price they would demand to forgo the option to draw a single number from the distribution. So we could assess possible perceptual biases, participants also completed a second task where they had to estimate the means of the same distributions. In this second task, accuracy was incentivized with respect to how closely the estimates matched the theoretical mean.

For both tasks, a single trial consisted of a rectangular box presented on the computer screen representing a distribution to draw from and a smaller gray box displayed below indicating where participants could type in their answers (see Figure 2 for a schematic). Participants could sample freely from the given distribution by pressing <space>, which was followed by a number presentation for 250 ms in the middle of the larger box. After the number disappeared from the screen, an additional sample could be drawn. Each presented number was generated as a random draw from the respective underlying frequency distribution, rounded to its nearest integer. After having drawn at least one sample, participants could enter their answer into the gray box. Sampling was also possible after entering a number and the number could be revised. To end a trial, a number had to be typed in and had to be confirmed with <enter>.

Outcome Distributions. We constructed 24 continuous number distributions by combining four different means (80, 100, 130, 160), two standard deviations (5, 10), and three shapes (normally distributed, 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. We used four levels of mean mainly to make the different sequences' means noticeable, in an effort to keep participants engaged in the task and hence to increase the number of trials. The different distributions were presented in randomized order and were the same in both the valuation and the estimation task.

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 participants' own pace. Each participant completed two blocks consisting of 24 trials each. In one block they had to estimate the mean of the number sequences and in the other they had to report their certainty equivalent. Block order was counterbalanced between participants.

Payment was determined by randomly selecting one answer across both blocks. If the trial was in the WTA block, a BDM procedure was implemented (Becker, DeGroot, & Marschak, 1964): A random number was uniformly drawn between 0 and the theoretical mean of a given distribution. When the random number was below the participant's answer for this trial, then the 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 absolute difference of the estimate and the true mean was subtracted from the true mean. In a final step, the obtained points were exchanged into Australian dollars (AUD) with a 20:1 ratio and paid out in cash.

Method Experiment 2

Experiment 2 was a direct, preregistered replication of the first experiment (https://osf.io/ehkuz/). The only difference was a change in participants' instructions. Anecdotal interviews of participants in the first study indicated some difficulties in comprehending the incentive scheme (particularly the BDM auction). Hence, in the second study 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 on an extra button on the screen. About one-third of participants made use of this option in each block.

Participants and Data Analysis Experiments 1 and 2. Because both experiments used the same stimuli and procedure, we included a study dummy variable across all statistical analyses. It never came out significant and hence we pooled both data sets to increase statistical power. Furthermore, there were no differences between participants who read the incentive schemes and those who did not in the second experiment.

We tested 53 participants in the first and 58 participants in the second experiment. Sample size was chosen based on the availability of a convenience sample 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). Participants' age and sex were not assessed, but in the subject pool the mean age was 19 years and approximately 70% were women.

Prior to analyzing the data we excluded answers further away than 5 standard deviations from the distribution's mean (21 trials in the first and 33 trials in the second experiment out of 5,232 total trials across both experiments). Further, two participants from the first experiment were excluded for not complying with the task: One participant sampled only once in each trial (the minimum to continue) and another participant gave only two answers within 5 standard deviations from the true mean.

This left us with 109 participants.

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. In particular, variance and skewness were dummy coded and the mean was treated as a continuous predictor variable. As dependent variables we defined the logarithm of sample size and participants' accuracy, quantified as the deviation of their answers proportional to the distributions' true mean. The last measure is similar to the (exponential) signed order of magnitude error that is sometimes reported in the literature (Brown & Siegler, 1992). Regression analyses that used the characteristics of the actually experienced samples rather than the theoretical characteristics as independent variables led to qualitatively similar results. For the second experiment, all exclusion criteria and the statistical regression analyses were preregistered.

Results

On average, participants drew $M = 28.81$ samples from each distribution ($Mdn = 21$, $SD = 31.25$). There was no difference in sample size between task types (estimation: $M = 28.51$, $Mdn = 21$, $SD = 28.79$, valuation: $M = 29.12$, $Mdn = 21$, $SD = 33.54$), $t(108) = 0.20$, $p > .250$. Table 1 shows the regression results with the logarithm of the number of drawn samples as dependent variable. As can be seen from the table, only variance had an effect on sample size: The higher the variance, the more participants sampled. This is in line with previous findings in the literature (Ashby, 2017; Lindskog, Winman, & Juslin, 2013) and it is adaptive in the sense that more samples mitigate higher uncertainty.

Valuation Task. Figure 3 (left) plots the proportional deviation of participants' answers in the valuation task from the true means across the different experimental conditions. As can be seen in the figure, participants gave lower certainty equivalents than the theoretical (i.e. true) means. The average deviation from the true mean was $M = -4.78$ ($Mdn = -3.13$, $SD = 16.45$). This is corroborated by a t test showing that certainty equivalents were significantly lower than the theoretical means, $t(108) = -4.65$, $p < .001$.

Higher variance led to lower certainty equivalents compared to lower variance sequences ($M = -4.68$, $Mdn = -3.13$, $SD = 19.18$). The middle column of Table 2 shows the regression results for the valuation task. In particular, the parameter for variance is negative (-4.73, $SE = 1.20$)—that is, higher variance led to significantly lower valuations. Together with the result of overall undervaluation of the mean, these results are consistent with the notion of risk-averse preferences.

Skewness also has a significant effect on economic valuations. Participants gave lower values to left- compared to right-skewed distributions with a mean difference between these two distributional forms of $M = -5.40$ ($Mdn = -1.52$, $SD = 18.41$). In line with this, the regression analysis shows significant effects, indicating that left-skewed outcome distributions were valued lower than normally distributed ones (-2.36, $SE = 0.63$) and that right-skewed outcome distributions were valued higher than normally distributed ones (3.02, $SE = 0.60$). Thus participants had a preference for right-skewed distributed outcomes.

Finally, results indicate that the mean had a significant positive effect (0.03, $SE = 0.01$) on participants' valuations. In particular, the proportional deviation from the theoretical mean got smaller as the theoretical mean increased from 80 to 160. Supposedly, this is because the variability relative to the mean was lower in trials with a mean of 160 than in those with a mean of 80. This is a consequence of the design choice to hold the absolute variance constant across different mean levels. Hence, taking the absolute deviation of the answers from the theoretical mean as dependent variable, there was no significant effect of the mean, whereas all other predictors remained significant.

Estimation Task. The mean estimates within each condition are depicted in Figure 3 (right). Like in the valuation task, participants underestimated the theoretical mean of the number sequences across all conditions ($M = -1.39$, $Mdn = 0$, $SD = 9.57$).

A t test revealed this underestimation to be significant, $t(108) = -3.00$, $p = .003$.

Underestimation was more pronounced for sequences with high variance compared to those with low variance ($M = -0.83$, $Mdn = 0$, $SD = 12.27$). Table 2 (right column) shows the respective regression results. As can be seen from the table, there is a significant effect of variance on estimation deviation in the direction descriptively observed (-1.33, $SE = 0.59$). Together with the effect of overall underestimation, this effect is in accordance with a compressed mental number line.

Furthermore, mean estimates for left-skewed distributions were lower than for right-skewed distributions ($M = -2.91$, $Mdn = -1$, $SD = 12.20$). The regression results reveal that mean estimates of right-skewed distributions were significantly higher (1.79, $SE = 0.47$) and mean estimates of left-skewed distributions significantly lower (-1.18, $SE = 0.47$) than mean estimates of normally distributed sequences. This is consistent with the idea of number compression and with overweighting of extreme outcomes.

Finally, the proportional deviation from the theoretical mean got smaller with higher means (0.01, $SE = 0.01$). As in the valuation task, this effect most likely was due to a decrease in relative variability as the mean increased. Inserting the absolute deviation from the theoretical mean as dependent variable shows no effect of the mean level on the deviation, whereas the effects of variance and skewness are robust to this change.

Comparing Estimation and Valuation. As described above, we found qualitatively similar answer patterns in the estimation and valuation tasks. The observed effects were smaller for the estimation task though. The observed overall ratio of underestimation to undervaluation was .29. This ratio can be interpreted as the relative influence of cognitive biases on valuation. Taking the difference between low- and high-variance trials separately for the two tasks and calculating the ratio of these two differences results in .18. Finally, taking the difference between left- and right-skewed trials separately for the two tasks and calculating the ratio of these differences gives .54. Together, these results suggest that participants' responses in the valuation task could be partly attributed to basic perceptual and cognitive regularities in their number estimation, particularly so for the effect of skewness.

Discussion Experiment 1 and 2

Experiment 1 and 2 together showed robust effects of overall undervaluation as well as effects of the variance and the skewness that are in line with behavior in other contexts of risky choice. In the estimation task, similar effects were observed which is evidence that answers in the economic valuation can partly be explained by the process of number integration. Underestimation, the effect of variance, and higher estimates for right- compared to left-skewed distributions were consistent with the compressed mental number line. Overweighting of extremely high numbers is consistent with higher estimates for right- compared to left-skewed distributions, but would have predicted higher estimates for high variance compared to low variance distributions. There was no evidence for underweighting of rare events in our data. Finally, increasing the mean led to less undervaluation and underestimation, which was not in line with any of the reviewed theories above. As a limitation, the mean effect vanished when using as dependent variable the absolute difference between the valuation and the theoretical mean compared to using the proportional difference. As noted, a possible confound was the variance, which was the same for all mean levels. Thus there was proportionally less variability for high compared to low mean sequences (see also Whalen et al., 1999).

Experiment 3

To overcome this limitation of the previous studies, we conducted a new experiment. In this experiment we held the variability constant across means and we increased the range of means to have more power to detect an effect of the mean on participants' answers.

Method

Material. The third Experiment uses the same design and task structure as the previous ones: Again, participants sampled to learn about number distributions as often as they wanted. In a within-subject design, participants had to estimate the mean in 24 trials and gave their certainty equivalent in 24 trials each. The different tasks were pooled in two distinct blocks with randomized trial order. Incentivization was changed for the certainty equivalence task compared to the first two experiments,in that a bid was selected from zero to the 99% quantile of the respective distribution. To check for an effect of number compression more specifically, we changed the number distributions by holding the relative variability (variance/mean) constant across different mean levels. Furthermore, we increased the number of variability levels from 2 to 3 (at levels of 5, 10, and 20 compared to a mean of 100) and we introduced a stronger mean manipulation using 8 different levels: 30, 50, 75, 100, 130, 160, 200, and 250. In order not to inflate the number of trials too much, we omitted the skewness manipulation.

Participants & Procedure. We recruited 120 participants from the University of Geneva subject pool. The sample size was determined prior to the start of the experiment, based on the results in the previous experiments. The experiment was conducted on single computer work stations. The experiment lasted on average about 30 minutes. The participants average age was 23 years (Median = 22, SD = 5.65). 40 participants were male and 80 female. On top of the show-up fee of 20 CHF, the average choice dependent bonus was 122.90 points (approx. 6 CHF; Median = 121.96, range = [21.85; 365.26].

Results

On average, participants drew $M = 28.47$ samples from each distribution ($Mdn = 20$, $SD = 29.22$). There was no difference in sample size between task types (estimation: $M = 29.53$, $Mdn = 20$, $SD = 29.53$, valuation: $M = 27.39$, $Mdn = 20$, $SD = 28.87$), $t(119) = -1.15$, $p > .250$. Table 3 shows the regression results with the logarithm of the number of drawn samples as dependent variable. As can be seen from the table, the mean significantly affects the sample size. The higher the stakes in a given trial, the higher is the sample size. In addition, variance had an effect on sample size: The higher the variance, the more participants sampled similar to the effect in

Experiments 1 and 2. Finally, the coefficient for task type is significant in this regression, revealing that participants sampled slightly more in the estimation compared to the valuation task.

Valuation Task. Figure 4 (left) plots the proportional deviation of participants' answers in the valuation task from the true means across the different experimental conditions. As can be seen in the figure, participants gave lower certainty equivalents than the theoretical (i.e. true) means. The average deviation from the true mean was $M = -5.55$ ($Mdn = -3$, $SD = 14.74$). This is corroborated by a t test showing that certainty equivalents were significantly lower than the theoretical means, $t(119) = -3.39$, $p < .001$.

Higher variability relative to the mean led to lower certainty equivalents compared to sequences with lower variability (variability 5: $M = -1.72$, $Mdn = 0$, $SD = 4.97$; variability 10: $M = -3.38$, $Mdn = -1.33$, $SD = 10.04$; variability 20: $M = -5.54$, $Mdn = -3$, $SD = 17.77$). The middle column of Table 4 shows the regression results for the valuation task. In particular, the parameter for variance is negative for both higher levels of variability (-2.01, $SE = 0.48$, -5.42, $SE = 0.48$). Also the effect is larger for the high compared to the medium variability (both in contrast to the low variability trials).

Since we explicitly manipulated variability, we can now differentiate the effect of variability from the effect of higher numbers. Results indicate that there is no significant effect of the mean level ($b = -0.0006$, $SE = 0.003$) on participants' valuations. In particular, we found no evidence for more risk-aversion for higher mean sequences.

Estimation Task. The mean estimates within each condition are depicted in Figure 4 (right). Similar to the previous experiments, participants underestimated the theoretical mean of the number sequences across all conditions ($M = -1.62$, $Mdn = 0$, $SD = 8.88$). A t test revealed this underestimation to be significant, $t(119) = -3.61$, $p < .001$.

Higher variability relative to the mean led to lower estimates compared to sequences with lower variability (variability 5: $M = -0.64$, $Mdn = 0$, $SD = 3.58$;

variability 10: $M = -1.69$, $Mdn = 0$, $SD = 8.19$; variability 20: $M = -2.51$, $Mdn = 0$, $SD = 12.38$). The right column of Table 4 shows the regression results for the estimation task. In particular, the parameter for variance is negative for both higher levels of variability (-0.79, $SE = 0.53$, -1.43, $SE = 0.54$). Also the effect is larger for the high compared to the medium variability (both in contrast to the low variability trials).

As in the valuation task, the proportional deviation from the theoretical mean was not affected by higher means (0.0005, $SE = 0.002$). This suggests that the observed numbers were not subject to a compressed mental number line.

Comparing Estimation and Valuation. As in the previous experiments, we found qualitatively similar effects of variability and mean on valuation and estimation answers. To quantify the influence of basic cognition on economic valuations, we again calculated the respective ratios of under*estimation* to under*valuation*.

The observed overall ratio of underestimation to undervaluation was again .29. Taking the difference between low- and high-variability trials separately for the two tasks and calculating the ratio of these two differences results in .26. Together, these results confirm that participants' risk-averse responses in the valuation task could be partly attributed to systematic biases in their number estimation. Finally, we divided trials into low-mean vs. high-mean trials and calculated the ratio of differences between these two task types between estimation and valuation, which results in a value of 0.05. The effect of the mean is thus much stronger in valuation compared to estimation, but note that the effect of the mean was not significant in either the regressions on valuation or on estimation trials.

A new Sample Weighting Model

As seen above, the behavioral results of the estimation and valuation of continuous outcome distributions are not fully consistent with either of the above revised theories. In particular, no effect of the mean on estimation and valuation is not in line with the compressed mental number line. On the other hand, underestimation and -valuation as well as stronger underestimation and -valuation for high compared to low variance distributions is not in line with the overweighting of high numbers.

Here, we present a post-hoc model of sample weighting that can account for the qualitative data pattern. A closer look into the literature reveals that behavior characterized as if people overweight high numbers was mostly found when people have to choose the preferred/ higher number stream of two options (Glickman, Tsetsos, & Usher, 2018; Ludvig & Spetch, 2011; Spitzer, Waschke, & Summerfield, 2017; Tsetsos et al., 2012). Interestingly, when people estimated whether the mean of a number sequence was lower or higher than a certain amount, people also overweighted low numbers (Experiment 5 in Kunar, Watson, Tsetsos, & Chater, 2017). The authors conjectured that the overweighting thus depends on the participants' goals in a task. Presumably, for estimation or valuation tasks, both very low and very high numbers capture attention. More concretely, the results in our experiments are consistent with the idea that people overweight small numbers more than large numbers in a given sequence of numbers. Mathematically, small and high numbers in a sequence can be defined by their distance to the median of the sequence. In the following we describe it as the proportional absolute deviation (pad) from the median:

$$pad = \frac{abs(sample - median(sample))}{median(sample)} + 1. \tag{1}$$

Here, the median is taken from the given distribution. The absolute deviation between sample and median guarantees the same scale for numbers below and above the median. Dividing by the median results in the percentage deviation, which then leads to the same weighting of percentage deviations independent of the mean. This is in line with many psychopyhsical approaches of mapping objective entities to subjective perceptions in general (Stevens, 1957) and in number perception and representation (e.g. Izard & Dehaene, 2008; Whalen et al., 1999). We add 1 to the ratio to have a more natural interpretation: Samples that are exactly the median have a weight of 1 and all samples deviating from the median have a higher weighting with the weighting increasing with absolute distance. In a final step all weights are divided by the sum of weights to result in weights that sum up to one. To account for individual heterogeneity and for different weighting of samples below and above the median, we introduce two parameters:

$$w(sample) = \begin{cases} pad^{\alpha}, & \text{if } sample - median(sample) \leq 0 \\ pad^{\beta}, & \text{otherwise.} \end{cases} \quad (2)$$

For $\alpha, \beta > 0$, samples away from the median get more weight than samples equal to the median. Assuming $\alpha > \beta$ means that samples below the median get more weight than samples above the median. As an illustration, Figure 5 shows the weightings of 1000 samples from a normal distribution with mean (and median) equal to 100 and standard deviation of 20. In this example $\alpha = 4$ and $\beta = 0.5$. This overweighting of low compared to high numbers directly leads to the underestimation or -valuation of the mean. Quantitatively, the smaller the parameter, the flatter the curve and thus the lesser overweighting of extreme numbers. That way, the different levels of underestimation compared to undervaluation, as observed in our studies, could be modeled by differences in the ratio between α and β: When α and β are the same, the predicted mean of a symmetric number sequence is equal to the true mean. The larger α is compared to β the more underestimation or -valuation is predicted. In case of asymmetric, that is right- and left-skewed distributions, the absolute level of the parameters plays a role. The larger the absolute values of α, β, the stronger the difference in underestimation or -valuation between left- and right-skewed distributions.

As numerical examples, we now construct distributions with different characteristics mimicking the stimulus material in the experiments. We sample from them 10,000 times and use again the parameter values $\alpha = 4$ and $\beta = 0.5$. This model predicts an answer of 99.62 for a low variance normal distribution ($m = 100$, $sd = 5$) compared to 92.09 for a high variance normal distribution ($m = 100$, $sd = 20$). This is in line with stronger underestimation for high compared to low variance distributions in our experiments. Doubling the mean of the distribution and keeping the relative variability constant ($m = 200$, $sd = 40$) leads to a predicted answer of 184.07. Comparing the percentage deviation from the true mean of -7.96% with the same value for the lower mean distribution of -7.91% shows that the mean does not strongly influence the deviation. This was the main finding in Study 3. A right-skewed distribution ($m = 100$, $sd = 20$, $skew = 2$) with most samples (slightly) below the mean and some samples highly above the mean is predicted to lead to comparatively little deviation of 98.74. In contrast, a left-skewed distribution ($m = 100$, $sd = 20$, $skew = -2$) with most samples (slightly) above the mean, but some extremely low samples, is predicted to lead to a much stronger deviation of 79.86. This reproduces the empirically found effect of skewness on answers in our experiments.

This model should illustrate how the effects found in our experiments can be accounted for in a simple mathematical model. It was built on frequently used concepts in the numeric cognition literature, like proportional deviations from the median and overweighting of extreme events. Whereas the model provides a parsimonious explanation for the observed patterns, it was constructed post-hoc after seeing the data and thus awaits further validation and rigorous testing in future research. Also, even though the assumed cognitive processes have some plausibility, the model assumes the median of the sequence of samples to be known beforehand. This was not true in our experiments. Instead people most likely have prior expectations about the range of outcomes and might not keep consciously track of the median with every sample. Thus, using the median in the model is just an approximation of a more complex process. Finally, the model needs many samples to lead to stable predictions. Thus, when humans are sampling voluntarily, the amount of samples - although leading to quite accurate summary statistics - leads to a lot of variability on the level of samples, which are the inputs to this model.

Discussion

In three experiments, participants sampled number sequences and either estimated their means or stated their willingness to pay for drawing an uncertain outcome from the respective distributions. Results indicate qualitatively similar answers in both tasks but also show clear quantitative differences, suggesting that economic preferences can be partly explained by cognitive biases in perceiving and integrating numeric information.

Overall, participants underestimated the mean of a sequence of numbers. A direct comparison to economic valuations, where people gave certainty equivalents below the distributions' actual means, showed that about one-third of this undervaluation can be attributed to an estimation bias. Underestimation and undervaluation were stronger for sequences with high variance. Here, about one-fourth of the undervaluation could be attributed to an underestimation bias. These patterns are in line with the compressed mental number line and were not predicted by underweighting rare events or by overweighting extremely high events. In addition, underestimation and undervaluation were more pronounced for left-skewed over right-skewed distributions. Here, about half of the variance in the observed undervaluation could be attributed to underestimation. This pattern is again in line with the compressed mental number line, but was also predicted by an overweighting of high numbers. Finally, when keeping variability constant relative to the mean as in Experiment 3, estimation and valuations did not vary significantly with the sequence's mean. This can be accommodated for by assuming only local (i.e. trial by trial) effects of extremeness or rarity, but are not in line with the compressed mental number line.

The Cognitive Process of Economic Valuation

Biased estimations of number sequences can be explained by introducing the idea of an intuitive (nonverbal) number sense that guides the perception and integration of numbers (Brezis et al., 2015; Feigenson et al., 2004; Gallistel & Gelman, 2000). To capture the cognitive processes that govern overall magnitude judgments from sequences of symbolic numbers, we referred to two different theoretical frameworks: First, a compression of the numeric scale and thus a concave psychophysical mapping of objective numbers to subjective numerosity. Second, systematic differences in the weighting of single samples, either through underweighting of rare events (Wulff et al., 2018) or overweighting of extreme events (Ludvig & Spetch, 2011).

A compressed mental number line can explain the overall underestimation and the stronger underestimation of high-variance sequences that we observed. Furthermore, it is consistent with higher mean estimates for right-skewed compared to left-skewed distributions. Our experiments show that this number sense can also (partly) explain economic valuations from experience. To the contrary, the notion of a compressed mental number line is inconsistent with our finding that underestimation did not intensify for higher means (controlling for relative variability). The similar concept that risk-aversion (i.e. compression in a power utility function) increases as the stakes become larger seems to be intuitively plausible though and has also been found in economic gamble tasks where the mean was increased by a factor of up to 90 (Holt & Laury, 2002). In contrast to this, the largest difference in our experiments was less than 10 times (i.e. 30 vs. 250). Thus, perhaps a mean effect can be found with a stronger manipulation of the mean.

Overweighting of extreme events is also consistent with our finding that people estimated higher means for right- compared to left-skewed distributions. Such an overweighting pattern might occur through attention and memory effects that render extreme outcomes easier to memorize and to retrieve (Kahneman et al., 1993; Madan et al., 2014; Parducci et al., 1968). Our empirical evidence supports such a weighting and identified it as an important source of the observed skewness preference in the economic valuation task. In our experimental design, extreme outcomes are always also rare outcomes. Thus we cannot rule out that the overweighting of rare events drove the skewness effect on estimation and valuation. Yet, in a recent study, Ludvig, Madan, McMillan, Xu, and Spetch (2018) found evidence for a stronger effect of the extremeness than the rarity of a numbers in a sequence.

On the contrary, the hypothesis that extreme numbers are overweighted did not predict overall underestimation of the mean and stronger underestimation for higher variance sequences. In addition, interpreting extremeness as the highest number predicts higher estimates as the variance of the observed sequence increases. This is the opposite of our finding. Higher estimates for sequences with larger variance would lead to risk-seeking behavior, which has indeed been found in a previous study where people chose between outcome sequences with different variance (Glickman et al., 2018; Ludvig & Spetch, 2011; Spitzer et al., 2017; Tsetsos et al., 2012). This suggests that the way samples are subjectively weighted differs depending on whether the task requires an explicit valuation or a choice between different options. Valuation and choice are qualitatively different indeed: Whereas in binary choice tasks two number streams have to be processed and integrated, only one number stream had to be heeded in a valuation task. Additionally, the higher level goals of the decider might differ in a valuation compared to a choice task (Kunar et al., 2017).

Finally, our data are not consistent with the hypothesis that rare events are underweighted. Underweighting of rare events could neither explain overall underestimation, nor the effect of variance. The effect of skewness was predicted to go into the opposite direction. Our results appear to be at odds with the more typical underweighting of rare events previously reported in the DFE literature (Wulff et al., 2018). However, our stimuli were continuous and thus generalizing the notion of rarity in number samples. We conclude that underweighting of rare events does not prevail in the context of continuous distributions. Even in binary outcome distributions, recent findings demonstrate seeming overweighting of rare events in DFE tasks examining choices between two two-outcome distributions (Glöckner et al., 2016; Kellen et al., 2016). Finally, some of the above theories of the imprecise number sense were mainly developed for the processing of non-symbolic magnitudes. Future studies could examine the predictions of these theories for a stream of non-symbolic magnitudes (as for example in Dutilh & Rieskamp, 2016; Zeigenfuse, Pleskac, & Liu, 2014).

Towards an Integrative Cognitive Model

To capture all qualitative patterns of our data, we developed post-hoc a new model of sample weighting: The core feature of this weighting model is that in extension to previous models (e.g. Tsetsos et al., 2012), high as well as low numbers are overweighted compared to numbers close to the median. As discussed above, differences in the task (estimation/ valuation vs. choices for higher numbers as in Ludvig & Spetch, 2011; Tsetsos et al., 2012) could be an important moderator to the way numbers of a sequence are weighted. Assuming that numbers below the median are more strongly overweighted than numbers above the median is consistent with our findings (see above). In favor of such a model, past research reported evidence that attentional differences exist in the processing of numbers that signify losses compared to those signifying gains (Tom, Fox, Trepel, & Poldrack, 2007; Yechiam & Hochman, 2013). Assuming an internal reference point in our task that made people perceive numbers below the median as losses could thus explain why they receive more weights than numbers above the median. If such a loss frame is more pronounced in valuation than estimation, it can explain why high variance number distributions are more strongly undervalued than underestimated.

In sum, future research is needed to probe the presented model and to clarify the mechanism that leads to overweighting of high and low numbers. Another open task is to explain the cognitive processes that give rise to the differences in magnitude between estimation and valuation. Perhaps numeric cognition and preference formation processes are additive in that existing number integration biases are augmented in an economic context. But the relation could also be more complex, for instance, when both, numeric cognition and economic preferences mutually depend on the presentation format.

Another approach to think about the integration of numerical information is the theory of decision-by-sampling (Olivola & Sagara, 2009; Stewart, 2009; Stewart, Chater, & Brown, 2006). This framework assumes that the perception and cognitive processing of numbers depend on their rank in a reference distribution. Stewart et al. (2006) showed that distributions of monetary amounts in the gain and loss domain as well as the distribution of frequency words can lead to economic preferences similar to cumulative prospect theory (Tversky & Kahneman, 1992) such as risk aversion in the gain domain and overweighting of rare events. The valuation data we observed in our experiments is in line with prospect-theory predictions. Thus, decision-by-sampling theory could explain similar answer patterns in the estimation task. However, prospect theory with a concave or compressed utility function would also predict to find more undervaluation for higher mean sequences, which we did not observe.

As a final note, decision-by-sampling theory also assumes that the proximal distributions of numbers have an effect on answer patterns (Stewart, Reimers, & Harris, 2014; Walasek & Stewart, 2015). In the extreme case, when everyday experience with numbers would not affect behavior in the laboratory, a pure rank-based integration of numbers would not predict systematic underestimation for normally distributed outcome sequences as presented in our experiments. This is because the numerical differences between rank positions in the cumulative probability function of the normal distribution are symmetric at the low and high end of the distribution. Likewise, a pure rank-based number perception would also fail to account for our findings showing higher underestimation for high variance (compared to low variance) and stronger underestimation for left- (compared to right-) skewed distributions.

Differences in the Presentation Format and Estimation Biases

A widely studied format dependency in economic choice is the systematic difference in behavior between experience-based and description-based representations (i.e. the description–experience gap). In comparing choices between both formats, researchers found differences in the weighting of rare and extreme events (Hertwig et al., 2004; Madan et al., 2014). These differences were explained by assuming either undersampling of rare events or a recency bias that gives more weight to later samples (Wulff et al., 2018). Yet, recency has not been found consistently in the literature and the sampling bias is limited to certain outcome distributions. For example, a sampling bias cannot explain the format differences when choosing between 50–50 gambles (Madan et al., 2014). Another way to better understand the source of this format dependency could be to fit functional forms to the respective choice patterns. Yet, studies fitting cumulative prospect theory to both descriptive and experience-based choice data came to inconclusive results with respect to differences in the utility and probability weighting parameters (Abdellaoui et al., 2011; Glöckner et al., 2016).

Our results suggest another perspective on format dependencies: Presumably, the influence of numeric cognition is greater in experience-based tasks than in description-based tasks. This is the case because in the experience-based format often a large number of single samples (> 10) have to be processed and integrated sequentially. To the extent that the cognitive complexity of information integration differs between the two paradigms, it seems plausible that biases in basic number perception and integration contribute to the description–experience gap.

Incorporating the cognitive processes underlying number perception and integration can advance the predictive power of economic models that usually do not make predictions about concrete parameter values for utility and probability weighting functions depending on the format. For example, given that skewness valuations in our experiments highly depended on estimation bias, we would expect that skewness preferences would be less pronounced in choices from description where estimation errors presumably are smaller. In line with this reasoning, the effect of skewness on preferences in description-based choices is indeed mixed (Åstebro et al., 2015; Lichtenstein, 1965; Spiliopoulos & Hertwig, 2015; Taleb, 2004). Consequently, when modeling economic behavior, researchers should consider both basic cognitive and genuine preferential components. This distinction is particularly important when measuring preferences by comparing utility and probability weighting parameters across different task designs (see Tversky & Fox, 1995).

From an applied perspective, our results suggest that people will be more risk averse when it is hard to perceive and integrate the underlying numerical information. Our results further suggest that skewness preferences will depend on the perceptual salience of extreme events. This implies that economic decision making can be improved by improving people's estimates of outcome distributions. One way to facilitate information integration would be, for example, to present a list of all sampled outcomes in experience-based information acquisition (Kopsacheilis, 2017). Another application of this idea is research in empirical finance, where people have been found to invest in riskier, but also more profitable, assets when combining descriptive information with simulated experience of return sequences compared to a mere description condition (Bradbury, Hens, & Zeisberger, 2014; Kaufmann, Weber, & Haisley, 2013). This suggests that potential biases in the estimation of number sequences can be corrected by presenting descriptive information in addition to a number sequence. Moreover, there is evidence that investors' long-run return expectations of a company that is newly listed on the stock market (i.e. after an initial public offering) are positively skewed, which in turn can lead to losses because stocks are overbought on the first day (Green & Hwang, 2012). Given our finding of higher estimates of the means for right- compared to left-skewed distributions, one could train decision makers to give less weight to rare or extreme outcomes and be thus less susceptible to overbuying newly listed stocks.

Conclusion

The results of our experiments indicate that part of what is often framed as an economic preference may be the result of cognitive processes of number perception. Thus, researchers and practitioners alike would benefit from considering possible influences of number perception and integration on economic choices. This can help people to assess and execute preferences more reliably. Furthermore, it contributes to a better prediction of preferences in an economic context depending on the way the information is presented. Estimations of symbolic number sequences in our experiments are in line with a complex weighting scheme that overweights both very low and very high numbers with higher overweighting of low than high numbers. We think magnitude judgments of sequentially presented numbers are a good test for established models of numeric cognition. In addition, these and other models could help to understand format dependencies like valuation vs. choice and experience vs. description in economics from the perspective of numeric cognition.

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.

Ashby, N. J. (2017). Numeracy predicts preference consistency: Deliberative search heuristics increase choice consistency for choices from description and experience. *Judgment and Decision Making*, *12*(2), 128.

Åstebro, T., Mata, J., & Santos-Pinto, L. (2015). Skewness seeking: Risk loving, optimism or overweighting of small probabilities? *Theory and Decision*, *78*(2), 189–208.

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.

Barron, G. & Ursino, G. (2013). Underweighting rare events in experience based decisions: Beyond sample error. *Journal of Economic Psychology*, *39*, 278–286.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, *9*(3), 226–232.

Binswanger, H. P. (1980). Attitudes toward risk: Experimental measurement in rural india. *American journal of agricultural economics*, *62*(3), 395–407.

Bradbury, M. A., Hens, T., & Zeisberger, S. (2014). 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*.

Brown, N. R. & Siegler, R. S. (1992). The roleof availabilityinthe estimationof nationalpopulations. *Memory & Cognition*, *20*(4), 406–412.

Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. OUP USA.

Dutilh, G. & Rieskamp, J. (2016). Comparing perceptual and preferential decision making. *Psychonomic bulletin & review*, *23*(3), 723–737.

Erev, I., Glozman, 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.

Friedman, M. (1953). The methodology of positive economics.

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.

Golec, J. & Tamarkin, M. (1998). Bettors love skewness, not risk, at the horse track. *Journal of Political Economy*, *106*(1), 205–225.

Green, T. C. & Hwang, B.-H. (2012). Initial public offerings as lotteries: Skewness preference and first-day returns. *Management Science*, *58*(2), 432–444.

Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description–experience gap in risky choice: The role of sample size and experienced probabilities. *Journal of Behavioral Decision Making*, *21*(5), 493–518.

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.

Hey, J. D. (1995). Experimental investigations of errors in decision making under risk. *European Economic Review*, *39*(3), 633–640.

Holt, C. A. & Laury, S. K. (2002). Risk aversion and incentive effects. *American economic review*, *92*(5), 1644–1655.

Izard, V. & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, *106*(3), 1221–1247.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*(9), 697.

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.

Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 263–291.

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.

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.

Kopsacheilis, O. (2017). The role of information search and its influence on risk preferences. *Theory and Decision*, 1–29.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298.

Kraus, A. & Litzenberger, R. H. (1976). Skewness preference and the valuation of risk assets. *The Journal of Finance*, *31*(4), 1085–1100.

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*. R package version 2.0-32. Retrieved from https://CRAN.R-project.org/package=lmerTest

Lichtenstein, S. (1965). Bases for preferences among three-outcome bets. *Journal of Experimental Psychology*, *69*(2), 162.

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., Madan, C. R., McMillan, N., Xu, Y., & Spetch, M. L. (2018). Living near the edge: How extreme outcomes and their neighbors drive risky choice. *Journal of Experimental Psychology: General*.

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*, (just-accepted), 1–33.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, *7*(1), 77–91.

Menezes, C., Geiss, C., & Tressler, J. (1980). Increasing downside risk. *The American Economic Review*, *70*(5), 921–932.

Mosteller, F. & Nogee, P. (1951). An experimental measurement of utility. *Journal of Political Economy*, *59*.

Moyer, R. S. & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*.

Olivola, C. Y. & Sagara, N. (2009). Distributions of observed death tolls govern sensitivity to human fatalities. *Proceedings of the National Academy of Sciences*, *106*(52), 22151–22156.

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.

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica*, 122–136.

R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/

Rothschild, M. & Stiglitz, J. E. (1971). Increasing risk ii: Its economic consequences. *Journal of Economic Theory*, *3*(1), 66–84.

RStudio Team. (2015). *Rstudio: Integrated development environment for r*. RStudio, Inc. Boston, MA. Retrieved from http://www.rstudio.com/

Scheibehenne, B. (2017). The psychophysics of price perception: Evidence from the lab and from the field. *Decision*.

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. (1982). The expected utility model: Its variants, purposes, evidence and limitations. *Journal of Economic Literature*, 529–563.

Spiliopoulos, L. & Hertwig, R. (2015). Nonlinear decision weights or skewness preference? a model competition involving decisions from description and experience.

Spitzer, B., Waschke, L., & Summerfield, C. (2017). Selective overweighting of larger magnitudes during noisy numerical comparison. *Nature Human Behaviour*, *1*(8), 0145.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, *64*(3), 153.

Stewart, N. (2009). Decision by sampling: The role of the decision environment in risky choice. *The Quarterly Journal of Experimental Psychology*, *62*(6), 1041–1062.

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

Stewart, N., Reimers, S., & Harris, A. J. (2014). On the origin of utility, weighting, and discounting functions: How they get their shapes and how to change their shapes. *Management Science*, *61*(3), 687–705.

Taleb, N. N. (2004). Bleed or blowup: What does empirical psychology tell us about the preference for negative skewness. *Journal of Behavioral Finance*, *5*(1), 2–7.

Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. A. (2007). The neural basis of loss aversion in decision-making under risk. *Science*, *315*(5811), 515–518.

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

Tversky, A. & Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological Review*, *102*(2), 269.

Tversky, A. & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, *5*(4), 297–323.

Ungemach, C., Chater, N., & Stewart, N. (2009). Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)? *Psychological Science*, *20*(4), 473–479.

Walasek, L. & Stewart, N. (2015). How to make loss aversion disappear and reverse: Tests of the decision by sampling origin of loss aversion. *Journal of Experimental Psychology: General*, *144*(1), 7.

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.

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.

Yechiam, E. & Hochman, G. (2013). Loss-aversion or loss-attention: The impact of losses on cognitive performance. *Cognitive Psychology*, *66*(2), 212–231.

Zeigenfuse, M. D., Pleskac, T. J., & Liu, T. (2014). Rapid decisions from experience. *Cognition*, *131*(2), 181–194.

Table 1

Experiment 1 + 2: Sample Size in Valuation and Estimation

Distribution characteristic	Estimate
(Intercept)	2.95*** (0.08)
Mean	-0.0005* (0.0002)
Variance	0.11*** (0.02)
Right-skewed	-0.01 (0.02)
Left-skewed	-0.01 (0.02)
Valuation	0.01 (0.01)

Note. Effects of theoretical mean, variance, skewness, and task type on the log-transformed number of drawn samples (dependent variable) based on a mixed-effects regression with subject random intercepts. Standard errors in parentheses. *$p < .05$. **$p < .01$. ***$p < .001$.

Table 2

Experiments 1 + 2: Answers in Valuation and Estimation

Distribution characteristic	Valuation	Estimation
(Intercept)	-6.44*** (1.48)	-2.83** (0.86)
Mean	0.03*** (0.01)	0.01* (0.01)
Variance	-4.68*** (0.49)	-1.02** (0.35)
Right-skewed	3.04*** (0.60)	1.74*** (0.42)
Left-skewed	-2.38*** (0.60)	-1.19** (0.42)

Note. Effects of theoretical mean, variance, and skewness on percentage deviation of answers from the theoretical mean in economic valuation and estimation. All models with subject random intercepts. Standard errors in parentheses.

*$p < .05$. **$p < .01$. ***$p < .001$.

Table 3

Experiment 3: Sample Size in Valuation and Estimation

Distribution characteristic	Estimate
(Intercept)	2.790*** (0.0772)
Mean	0.0007*** (0.0001)
Variance10	0.0805*** (0.0156)
Variance20	0.0152*** (0.02)
Valuation	-0.0647*** (0.0127)

Note. Effects of theoretical mean, variance, and task type on the log-transformed number of drawn samples (dependent variable) based on a mixed-effects regression with subject random intercepts. Standard errors in parentheses.

*p < .05. **p < .01. ***p < .001.

Table 4

Experiment 3: Answers in Valuation and Estimation

Distribution characteristic	Valuation	Estimation
(Intercept)	-2.42 (1.28)	-1.21* (0.60)
Mean	0.0006 (0.003)	0.0005 (0.002)
Variability10	-2.01*** (0.48)	-0.79* (0.54)
Variability20	-5.43*** (0.48)	-1.43*** (0.54)

Note. Effects of theoretical mean and variance on percentage deviation of answers from the theoretical mean in economic valuation and estimation. All models with subject random intercepts. Standard errors in parentheses.

*$p < .05$. **$p < .01$. ***$p < .001$.

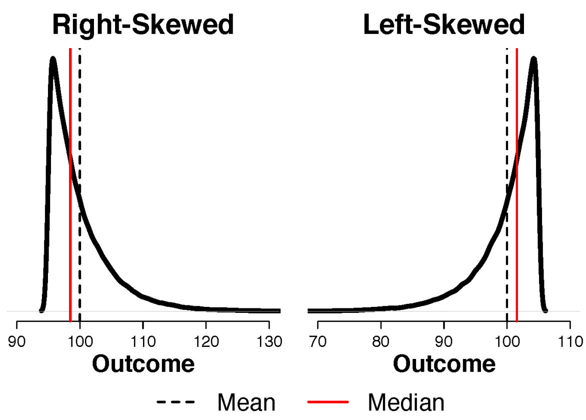

Figure 1. Right- and left-skewed distributions as used in the experiments with mean and median as vertical lines (in this example the mean is 100).

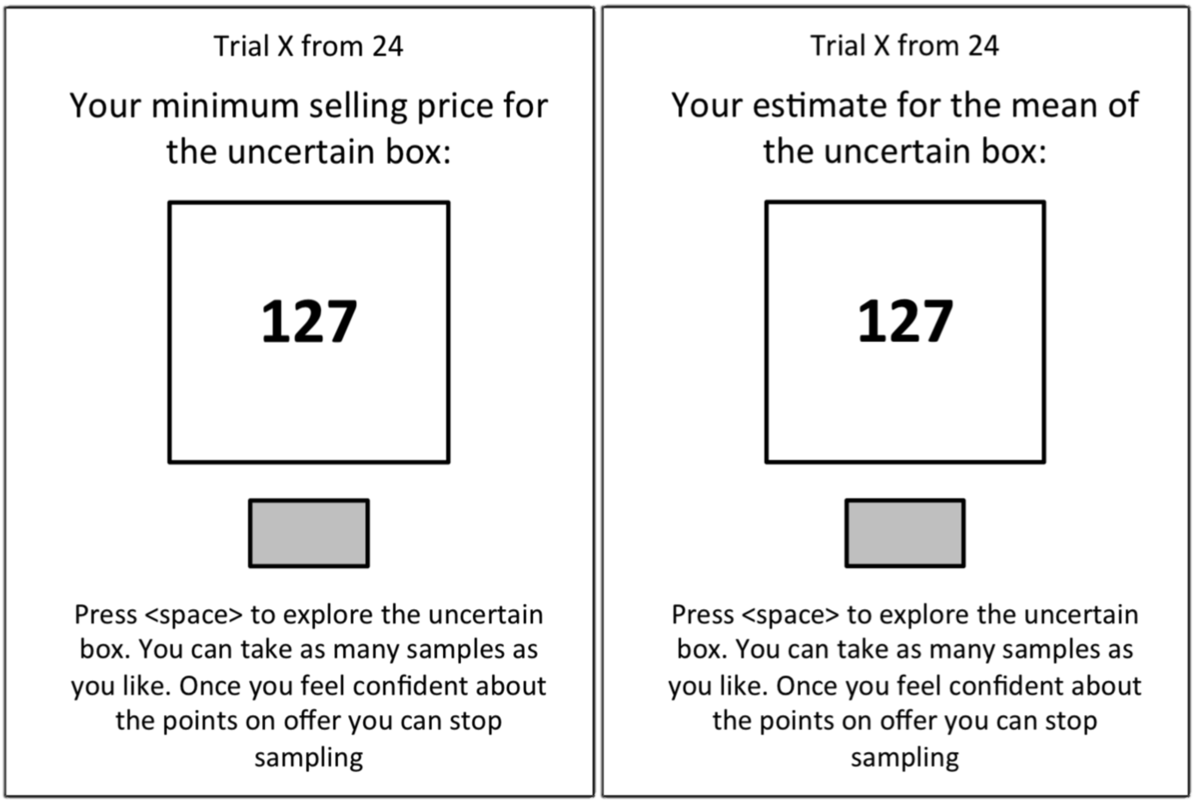

Figure 2. Schematic for one trial in the valuation (left) and estimation (right) task. Participants sampled from the white box and could type their answer into the gray box.

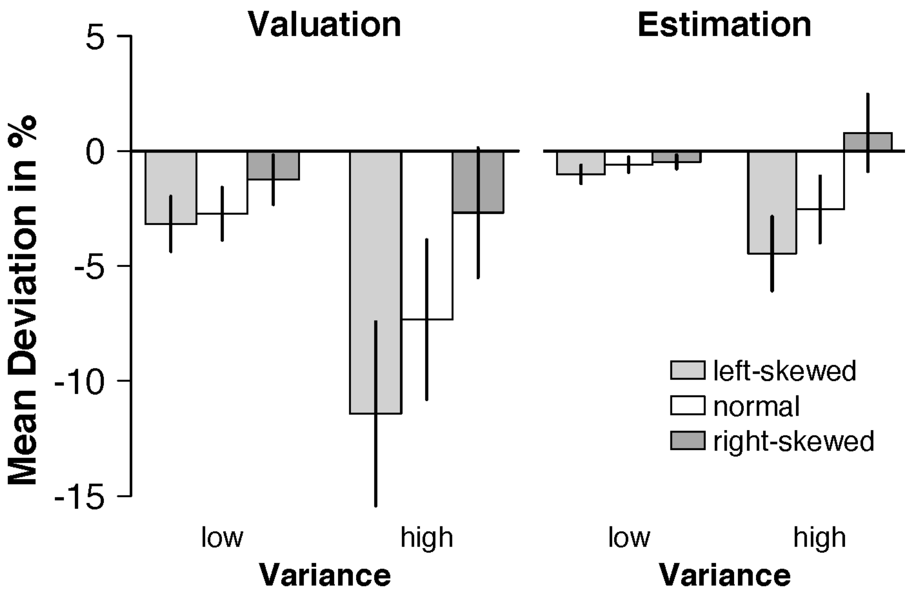

Figure 3. Answers in Experiments 1 and 2: The y axis shows percentage deviation of participants' answers from the distributions' theoretical means across different experimental conditions. Error bars are 95% confidence intervals.

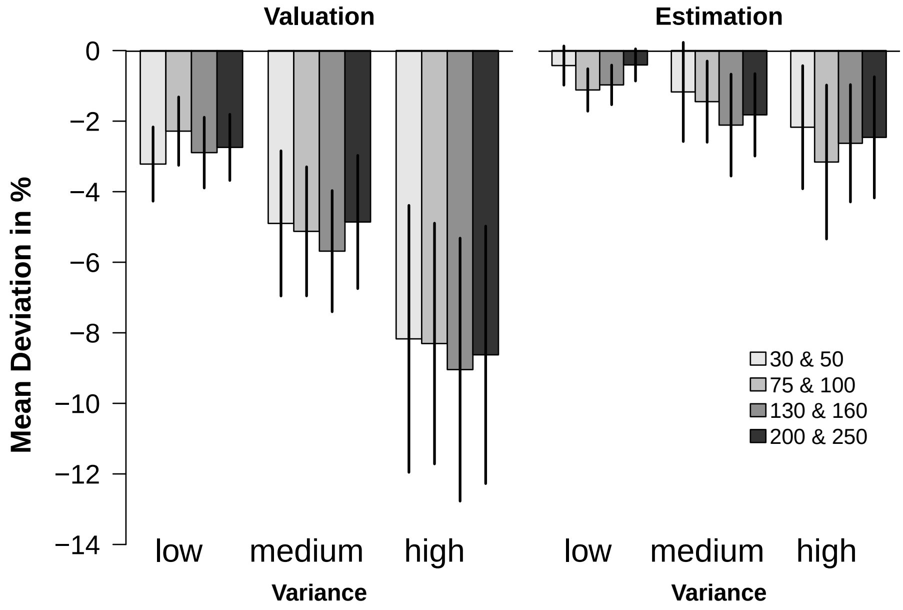

Figure 4. Answers in Experiment 3: The y axis shows percentage deviation of participants' answers from the distributions' theoretical means across different experimental conditions. Error bars are 95% confidence intervals.

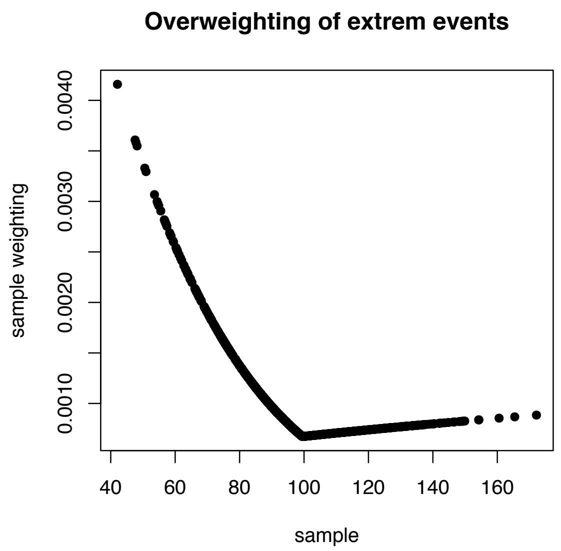

Figure 5. Weights of 1000 samples from a normal distribution with $mean = 100$ and $sd = 20$. Weights are calculated according to equation 1 and 2 with $\alpha = 4$ and $\beta = 0.5$.