

Learning Variability from Experience

Elizaveta Konovalova (konovalova@mpib-berlin.mpg.de)

Center for Adaptive Rationality,
Max Planck Institute for Human Development,

Gaël Le Mens (gael.le-mens@upf.edu)

Department of Economics and Business,
Universitat Pompeu Fabra, Barcelona, Spain

Abstract

Leading theories of risky choice predict that decision makers are sensitive to the variability of payoff distributions. Yet, little is known about how experience affects perceived variability. Existing empirical research on risky choice provides only inconclusive evidence about this issue because choices are not only affected by perceived variability but also perceived value and (unobserved) risk preferences. In re-analyses of experimental data and survey data from two nationally representative panels, we show that perceived variability strongly depends on sample variability. In a new experiment, we also demonstrate that perceived variability systematically depends on sample size, a result consistent with the predictions of a recent theoretical paper by the authors (Konovalova & Le Mens, 2017).

Keywords: Experience, Learning, Risky Choice, Sampling, Variability

Introduction

A large amount of research has studied decisions under risk. Leading models such as expected utility theory (Von Neumann & Morgenstern, 2007) or cumulative prospect theory (Tversky & Kahneman, 1992) assume that choice depends on the payoff distributions of the available alternatives as well as the risk preference of the decision maker. Two features of payoff distributions have received a particular amount of attention: the mean and the variance (the first and second moments). Consider, for example, expected utility theory. According to this theory, a risk-neutral decision maker will choose an alternative that maximizes expected payoff or, in other words, with the highest mean of its payoff distribution. And a risk-averse decision maker who faces a choice between several alternatives with the same mean payoff will choose an alternative with minimal payoff variance.

A number of researchers have noted that in many (maybe most) naturally occurring environments, decision-makers are not provided with explicit descriptions of the payoff distributions. Instead, they have to learn the relevant features of these payoff distributions from experience (Hertwig, Barron, Weber, & Erev, 2004; Weber, Shafir, & Blais, 2004). In such environments, to be valid, the choice theories which assume that payoff mean and variance affect risky choice have to invoke *perceived* mean and variance rather than *true* mean and variance.

It is well-known that when people learn about the mean of the payoff distribution of an alternative by sampling it a number of times, the perceived mean tends to be close to the sample mean (Busemeyer & Myung, 1992; Hogarth & Einhorn, 1992; Denrell, 2005; Le Mens, Kareev, & Avrahami, 2016).

Much less is known, however, about how experience affects perceived variance. In this paper, we address this issue.

With few exceptions, most of the research that bears to the question of how experience affects perceived variability comes from the literature on risky choice from experience (e.g., Hertwig et al., 2004; Weber et al., 2004). For example, in an influential paper, Weber et al. (2004) provided evidence that the coefficient of variation (the standard deviation divided by the mean) is a good predictor of risky choice in a setting where participants learned the payoff distributions from experience (Experiment 1). This suggests that both perceived mean and perceived variance matter in learning-by-sampling environments. The authors provided some evidence for the link between sample variability and perceived variability. Their study included settings that differed in terms of the variances of the payoff distribution, but with the same means. When the variance was higher, choices were more risk-averse. However, the investigators did not elicit perceived variance nor report the variance of the samples collected by participants. Because choice behavior is also affected by risk preferences and perceived value, the observed difference in behavior does not unambiguously indicate a difference in perceived variance.

The rest of the paper is made of two main sections. The first section focuses on the association between sample variance and perceived variance. We discuss existing evidence and report re-analyses of several existing datasets (an experiment and two datasets extracted from surveys of representative samples of the population of two countries). We find support for the hypothesis that perceived variance systematically depends on sample variance. We also find evidence that both sample variance and perceived variance tend to be lower than the true variance. These later results are consistent with the theoretical predictions of the model introduced by the authors in a recent paper (Konovalova & Le Mens, 2017). The second section concerns the sensitivity of perceived variance to sample size. Our review of the existing evidence and the results of a new experiment indicate that people tend to perceive larger samples as more variable than smaller samples. This is again consistent with the predictions of the model analyzed by Konovalova and Le Mens (2017). Moreover, consistent with the mechanism proposed by Konovalova and Le Mens (2017), we find that sample variance mediates this effect. By contrast to most prior research on decisions from experience, we do not rely on a task environment where decision makers

have a hedonic goal. Instead, we directly elicit beliefs about variance. This eliminates the potentially confounding effects of perceived mean and unobserved risk preferences.

Sample Variability and Perceived Variability

Existing Evidence

Early studies noted that central tendency and extreme observations were more salient than other observations (Hamilos & Pitz, 1977) and that participants put more weight on smaller deviations than on larger ones (Beach & Scopp, 1968).

The most comprehensive investigation of the association between sample variability and perceived variability was conducted by Kareev, Arnon, and Horwitz-Zeliger (2002). Their Experiment 1 was specifically designed to analyze this association. In this experiment, participants went through two tasks with the same structure (for brevity, we describe just one of the tasks): Participants first observed a population of 28 items that differed from each other on just one dimension. The items were paper cylinders of the same shapes colored up to a certain height. The height of coloring was the focal dimension. The coloring height was normally distributed with mean 6 cm and standard deviation 1.955 cm.

Participants then completed a comparison task: They were shown two additional populations of 28 items and were asked to identify the population most similar to the original. Unbeknown to the participants, one of the two comparison populations was the same as the population they saw. The other comparison population had higher variability (the distribution of coloring height had mean 6 cm and standard deviation 2.112 cm) or lower variability (same mean and standard deviation 1.811 cm).

The authors were interested in the proportion of participants who would select the non-identical population when it had higher or lower variability (the correct choice was to select the identical population). They found that when this alternative had lower variability, it was more likely to be selected than when it had higher variability. This result indicates that participants were sensitive to the variability of the population and that they had a systematic tendency to underestimate the variability.

Experiment 1 in Weber et al. (2004) also provides evidence that people are sensitive to sampled variability. These authors focused on risky choice situations where one of the alternatives had a sure payoff x , and the risky alternative had a probability p to yield a high payoff $y > x$ and $1 - p$ to yield a low payoff. Participants made choices based on experience: they were not provided with a description of the payoff distributions, but instead had to learn by sampling the two alternatives. The authors found that people were less likely to select the risky alternative when its coefficient of variation (CV) was high. This indicates that the perceived variability of the risky alternative was influenced by the sampled variability of that alternative.

These studies provide suggestive evidence that sample variability affects perceived variability. Yet, the evidence is

not as strong as it could be. In the study by Kareev et al. (2002), sample variability did not vary. And in the study by Weber et al. (2004), the samples observed by the participants were not analyzed. Next, we analyze data from a study that is not subject to these limitations.

New Analysis of Existing Experimental Data

We re-analyzed data collected by Goldstein and Rothschild (2014). In this online experiment, the authors told the participants that they had a very large bag with balls, that each ball had a number written on it, and that the range of numbers was 1 to 10. Participants were then shown 100 balls from the urn in a random order. It is important to note that the composition of the sample of 100 balls was not ‘random’, but was generated to be as close as possible to the generating distribution. In particular, the sample variance was essentially the same as the variance of the generating distribution. After seeing the sample, beliefs about the distribution of numbers were elicited using a tool designed by the authors called the ‘distribution’ builder. They also elicited the perceived 10%-90% range.¹ Goldstein and Rothschild (2014) write

After observing all 100 numbers from a randomly-assigned distribution and shuffle combination, respondents are told “Now imagine we throw the 100 balls you just saw back into the bag and mix them up. After that, we draw again 100 balls at random” [...] Respondents [were] asked “How many balls of each value (from 1 to 10) do you think we would draw?” By clicking on buttons beneath columns corresponding to the values from 1 to 10, respondents place 100 virtual balls in ten bins, ultimately creating a 100-unit histogram that should reflect their beliefs about a new sample drawn from the same population that gave rise to the sample they initially observed.

For the range task, participants were told that instead of 100 balls they would just draw one ball. Then they completed the following statements: “I am 90% certain the value of this ball would be greater than or equal to ...” and “I am 90% certain the value of this ball would be less than or equal to ...”. In a between participant design, the authors used six distributions (Figure 1) which had four unique variances.

Figure 2 shows that the variance of the elicited distribution is increasing in the variance of the sampled distribution. A one-way ANOVA shows that this effect is strongly significant ($F(3, 117) = 6.76, p < 0.001$). Similar but weaker results hold for the range task ($F(3, 116) = 2.50, p = 0.063$). These results indicate that perceived variability strongly depends on the variability of the sample. Next, we provide concurrent evidence from two non-experimental settings.

¹The distribution and range tasks were two of the tasks they used, they also similarly measured percentiles of the distribution. We focus on the data from these two tasks since it is the most comprehensive assessment of the distributional beliefs of the participants.

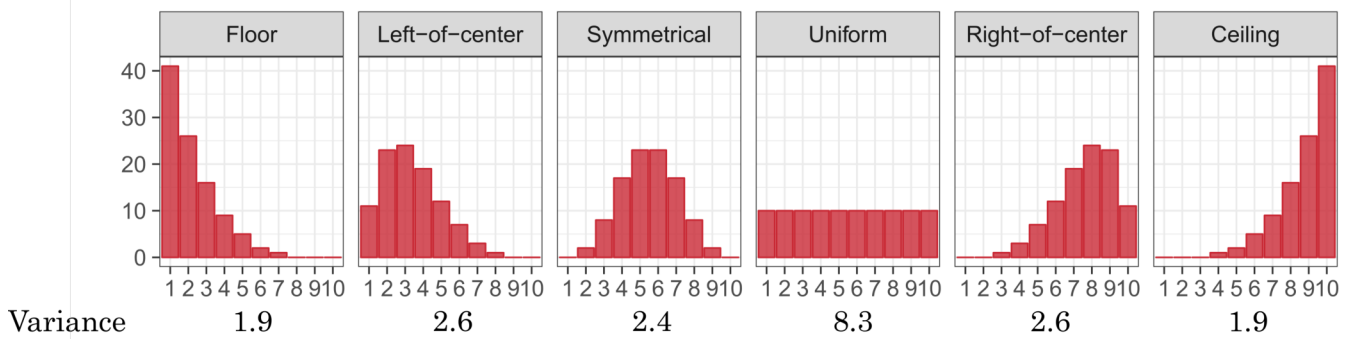


Figure 1: Distributions used in Goldstein and Rothschild (2014).

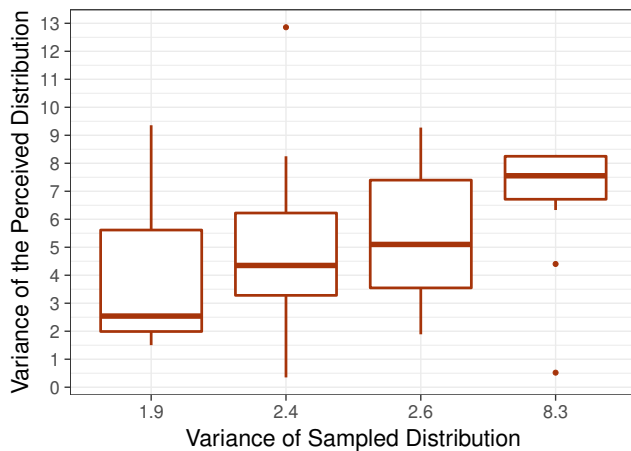


Figure 2: Analysis of the Goldstein and Rothschild (2014) data: Box plot of the impact of the higher variance of the sampled distribution on the variance of the perceived distribution.

Longitudinal Internet Studies for the Social Sciences (LISS) data

We analyzed data from the Longitudinal Internet Studies for the Social Sciences (LISS) panel. These data are from a representative sample of the Dutch population and were collected by CentERdata in collaboration with Galesic, Olsson, and Rieskamp (2012). The project explored the relationship between the social circles of the respondents (the individuals with whom they interact most frequently) and their perceptions of the national population as a whole. In this study, the authors asked respondents about ten characteristics related to their financial situation, friendships, health, work stress and education. The respondents reported their beliefs about the distribution of these characteristics on a 7-point scale. They were also asked to estimate the distribution in the general population of the Netherlands with questions such as “What percentage of adults living in The Netherlands fall into the following categories?”. In a second wave, participants were asked to provide the distribution in their social circle with questions such as “What percentage of your social contacts

fall into the following categories?”. ‘Social contacts’ were defined as “all adults you were in personal, face-to-face contact with at least twice this year.” (quoted from the codebook of the second wave of the study).

The authors were interested in how social sampling impacts beliefs about population characteristics. In their analyses, they assumed that available samples of the population were made of their social circles. Here, we rely on the same assumption.

We focus on one specific aspect of the social circle and perceived population distributions: their variance. For each of the ten characteristics, we regressed the variance of the perceived population distribution on the variance of the social circle distribution. The slope coefficient is significantly positive for all ten characteristics. It is also positive in a regression that pools the data about all ten characteristics and includes characteristic fixed effects (coefficient = 0.15, see Table 1). It is worth noting that the coefficients are somewhat far from 1. This indicates that the distribution in the social circle is not the only factor affecting the perceived population distribution. This is not surprising because it is unrealistic to expect that people’s only source of information about the population is their social circles. People interact with many others who are not part of their immediate social circles, read and watch about others in the media, etc. Yet, these results provide a clear indication that the variance of the sampled distribution affects the variance of the perceived population distribution.

Underestimation of True Variability In a recent paper, Konovalova and Le Mens (2017) demonstrated that for a number of measures of variability, sample variability (V_S) is more likely to be below the true variability than above the true variability (V_R): $P(V_S < V_R) > P(V_S > V_R)$. This prediction holds in particular for the case where the measure of variability is sample variance. We tested this prediction using the LISS panel data. For each characteristic, respondents were asked to indicate their position on the seven-level scale. This allowed us to construct the true population distribution. The data were collected in two waves. In each wave, participants indicated their position in the distribution. The results dis-

Table 1: Columns 1-2: Results of the regression analysis for the variance of the perceived distribution in LISS data. DV: Variance of the perceived population distribution (V_P); IV: Variance of the social circle (V_S). The results are shown for each domain (estimated constants are omitted) and for the whole panel with characteristic fixed effects. *** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$. Standard errors are in the parentheses. Columns 4-5: comparison between the variance of the perceived population distribution (V_P), the variance of the social circle distribution (V_S) and the variance of the real population distribution (V_R) in LISS Panel Data.

Characteristic	V_S	N	$P(V_P < V_R)$	$P(V_S < V_R)$
Amount of Stress	0.10*** (0.02)	1,407 (0.04)	.87	.96
Personal Income	0.12*** (0.02)	1,408 (0.03)	.52	.90
Household Income	0.13*** (0.03)	1,407 (0.02)	.72	.94
Wealth	0.12*** (0.03)	1,404 (0.03)	.70	.94
Number of Friends	.16*** (0.02)	1,408 (0.03)	.67	.93
Level of Education	0.16*** (0.03)	1,410 (0.02)	.26	.87
Number of Problems	0.19*** (0.04)	1,409 (0.02)	.72	.92
Number of Meetings	0.21*** (0.03)	1,299 (0.03)	.78	.93
Number of Conflicts	0.13*** (0.04)	1,087 (0.03)	.33	.88
Number of Dates	0.18*** (0.10)	277 (0.06)	.20	.74
Pooled data	0.15*** (0.01)	12,516 -	.61 -	.92 -

cussed in the text are based on a real distribution constructed from the responses of participants about their position collected in the first wave. The results are essentially the same for the distribution based on the second wave responses.²

Table 1 reports the proportion of respondents for which the social circle distribution had a variance lower than the variance of the true population distribution (see column $P(V_S < V_R)$). This was higher than 50% for all ten characteristics, as well as for the pooled data. A similar pattern was found regarding the proportion of respondents for which the perceived population distribution had a variance lower than the variance of the true population distribution (see column $P(V_P < V_R)$). This was higher than 50% for most characteristics. In the pooled data, the proportion of underestimation is 0.61. In summary, there is a general tendency for perceived variability

²There is a bit of irony in calling the ‘true population distribution’ a distribution constructed on the basis of a sample of smaller size than the true population (the population of the Netherlands). But because this sample is large, (about 1,400 people), its sample variance is very likely to be almost identical to the population variance.

ity to be lower to true variability, although the asymmetry is not as strong as for the sampled variability.

Sample Size and Perceived Variability

Konovalova and Le Mens (2017) also demonstrated that the underestimation tendency discussed at the end of the previous section systematically varies with sample size: It is strong when the sample size is small and milder when sample size increases. We formalize this in the following prediction: *the probability that sample variability is smaller than true variability is higher than chance and goes down with sample size*. Moreover, if, as we showed in the previous section, sample variability systematically affects perceived variability, then we should observe a similar tendency for perceived variability.

Existing Evidence

Existing evidence bearing on this prediction is limited. According to our literature search, the only published study that provides a direct test of this prediction is Experiment 2 in Kareev et al. (2002). Participants saw two populations of equal variance (this was unknown to the participants) then they were asked to indicate which of the two was the less variable. The stimuli were the same as in their Experiment 1 (discussed in an earlier section). Participants were asked to judge which of the two populations was more variable (on a unique dimension). Unbeknown to the participants, the two populations had the same distribution. They saw a sample from each population. For one population, participants saw the whole population (28 items). We call it the ‘large sample population’. For the other, they draw a random sample of 7 items. We call it the ‘small sample population.’ The majority of participants indicated the small sample population as the less variable. Participants also completed an incentivized task where the optimal choice was to select the less variable population (they were told that two items will be drawn from the selected population and that they would receive a bonus if they were close enough). Again, the majority of participants selected the small sample population. Overall, these results indicate that the participants perceived the small sample population as less variable than the large sample population.

Although this study provides evidence that the perceived variability of a distribution increases with sample size, an alternative explanation is possible. Without the information about the actual sample variability observed by the participants, it is not possible to rule out the hypothesis that the people perceive a large sample population as more variable even if the observed sample was not more variable. To address this limitation, we ran a new experiment.

Experiment

Design. Our design is inspired by features of the experiment in Goldstein and Rothschild (2014) and of Experiment 2 in Kareev et al. (2002). The flow of the experiment was as follows. After providing consent, participants received the

Table 2: Proportion of participants who indicated the large sample bag as the more variable. 95% Confidence intervals are in the brackets.

Question	All Observations	Conditional on Large Sample Bag Variance: $V_L > V_S$	Conditional on Large Sample Bag Variance: $V_L < V_S$	Difference in Proportions
Q1	.61 [.55, .67]	.68 [.61, .75]	.52 [.43, .60]	.17 [.05, .28]
Q2	.70 [.65, .75]	.79 [.72, .85]	.58 [.49, .67]	.21 [.10, .32]
Q3	.53 [.48, .59]	.64 [.56, .71]	.40 [.32, .49]	.24 [.12, .35]
# part.	303	173	130	–

following general instructions: “Imagine we have two extremely large bags: one with RED ping pong balls and one with BLUE ping pong balls. Each ball (both red and blue) has a value between 1 and 10 written on it. During the experiment, you will observe balls first from one bag and then from another. In the end, you will have to judge which bag has the larger variety of numbers on the balls.” Then, participants observed a random sample from one bag and in the following block a random sample from the other bag. The sample sizes were 5 and 50. The pairing of the color and the sample size was randomized as well as the order in which the two samples were presented. The samples were drawn from the same distribution. We used a symmetrical distribution which ranged from 2 to 9 with the following frequencies: [0.01, 0.06, 0.17, 0.26, 0.26, 0.17, 0.06, 0.01]. This distribution is a re-scaled and discretized beta distribution with parameters $\alpha = \beta = 5$.

Each participant observed a unique random sequence from the distribution. Before each sample, the participants saw a fixation cross for 450 milliseconds. Then digits appeared on the screen in quick succession (each digit remained on the screen for 600 milliseconds).

After participants observed the samples from the two bags, they answered three questions pertaining to the perceived variabilities of the two bags.

- **Q1:** This question was incentivized. Participants were told: “Suppose you select two balls from one of the two bags. Let us call A and B the numbers on the balls. Let D be the difference between these two numbers. You will get a bonus of D points. That is, the larger the difference between the two numbers, the higher your bonus (the bonus cannot be negative).” At the end of the experiment, two balls were randomly drawn from the chosen bag and participants were paid a bonus proportional to D. The goal was thus to select the bag with the higher variability.
- **Q2:** Participants were presented with a continuous slider where they indicated which bag had the larger “variety of numbers on the balls”. The minimal value of the slider was -100 (e.g., “The Red bag has more variety”). The maximal value was 100 (e.g., “The Blue bag has more variety”) and had a midpoint at 0 (e.g., “The Red and Blue bags have the

same variety”). (The colors at the end of the scales were randomized and the numeric values were not shown to the participants).

- **Q3:** Participants were asked to imagine they would pick two balls from each of the two bags. Then they were asked to indicate the bag for which they predicted the two numbers to be closer to each other.

Participants. We recruited 303 participants using Amazon Mechanical Turk. Participants received a fixed payment for their time and a bonus based on their responses to Q1.

Predictions. *Manipulation check:* We anticipated that for most participants the sample variability (variance) of the large sample bag (V_L^c) would be larger than the sample variability of the small sample bag (V_S^c): $P(V_L^c > V_S^c) > .5$. *Prediction about perceived variability:* Most participants will select the large sample bag as the more variable bag. *Prediction about the effect of sample variability:* The proportion of participants choosing the large sample bag will be higher when the large sample bag has the higher variability than when it has the lower variability.

Results. The results are consistent with our prediction. We report our analyses by using the corrected sample variance as the estimator of sample variability.

Manipulation check: For 57% of the participants, the sample variance of the large sample bag V_L^c was larger than the corrected sample variance of the small sample bag V_S^c : $P(V_L^c > V_S^c) = .57, 95\%CI = [.51, .63]$.

Sample size and perceived variability: Most participants perceived the large sample bag as more variable than the small sample bag. For Q1, 61% of the participants chose the bag of which they observed a larger sample. This proportion is significantly above 50% ($95\%CI = [.55, .67]$). For Q2, 70% of the participants selected a response on the scale that indicated that the large sample bag had “more variety” ($95\%CI = [.65, .75]$). The mean response was 33.33 ($95\%CI = [26.5, 40.2]$). This is significantly higher than the mid-point of 0. For Q3, 53% of the participants indicated that

balls from the bag of which they observed a smaller sample were closer to each other. This proportion is only marginally significantly different from 50% (95%CI = [.48, .59], $p = .13$).

Sample variability and perceived variability: We computed the proportion of participants who chose the large sample bag as the more variable when its sample variance was larger. For Q1 it is .68 (95%CI = [.61, .75], $n = 173$). The corresponding proportion conditional on the larger bag having the lower sample variance is .52 (95%CI = [.43, .6], $n = 130$). The difference in proportions is significantly higher than 0: $d = .17$, 95%CI = [.05, .28]. Similar results hold for Q2 and Q3 (see Table 2).

Summary Most participants perceived the large sample option as more variable than the small sample option even though the samples were generated from the same underlying distribution. Sample size had a positive effect on sample variability and the difference in sample variabilities had a positive effect on the difference in perceived variabilities. The tendency to perceive the large sample option as the more variable is thus at least partly explained by the difference in sample variabilities.

Discussion & Conclusion

Existing research acknowledges the importance of sample variance and size but implicitly assumes that the mapping between the sample and its mental representation is perfect (Osherson, Smith, Wilkie, Lopez, & Shafir, 1990; Konovalova & Le Mens, 2017). In this paper, we tested this assumption and provided direct (from experimental data) and indirect (from the analysis of survey data from a nationally representative panel of respondents) evidence of the relationship between sample variance and perceived variability. Additionally, our analysis shows that people's sample and perceived variance tends to underestimate the real variability. We also provided direct evidence that sample size has a positive effect on perceived variability and that this relation is at least partly mediated by sample variance.

We assumed that variance (sample variance or variance of the perceived distribution) is a psychologically relevant measure of variability. Our experiment provides suggestive evidence it is the case in at least some settings. Yet, existing research has shown that other measures of variability are sometimes more relevant. For example, Weber et al. (2004) convincingly argued that in risky choice situations, the coefficient of variation (CV) is a better measure of perceived variability than variance. Uncovering under what task environment sample variance, the coefficient of variation or other estimators of variability are the most relevant psychological constructs is an interesting avenue for future research.

Acknowledgements

Le Mens benefited from financial support from Spanish MINECO Grants PSI2013-41909-P and #AEI/FEDER UE-

PSI2016-75353, a Ramon y Cajal Fellowship (RYC-2014-15035), and a Grant IN[15]_EFG_ECO_2281 from the Fundacion BBVA. E. Konovalova was funded by Spanish MINECO Grant PSI2013-41909-P to G. Le Mens.

References

- Beach, L. R., & Scopp, T. S. (1968). Intuitive statistical inferences about variances. *Organizational Behavior and Human Performance*, 3(2), 109–123.
- Busemeyer, J. R., & Myung, I. J. (1992). An adaptive approach to human decision making: Learning theory, decision theory, and human performance. *Journal of Experimental Psychology: General*, 121(2), 177.
- Denrell, J. (2005). Why most people disapprove of me: experience sampling in impression formation. *Psychological review*, 112(4), 951.
- Galesic, M., Olsson, H., & Rieskamp, J. (2012). Social sampling explains apparent biases in judgments of social environments. *Psychological Science*, 0956797612445313.
- Goldstein, D. G., & Rothschild, D. (2014). Lay understanding of probability distributions. *Judgment and Decision Making*, 9(1), 1.
- Hamilos, C. A., & Pitz, G. F. (1977). The encoding and recognition of probabilistic information in a decision task. *Organizational Behavior and Human Performance*, 20(2), 184–202.
- 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.
- Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive psychology*, 24(1), 1–55.
- Kareev, Y., Arnon, S., & Horwitz-Zeliger, R. (2002). On the misperception of variability. *Journal of Experimental Psychology: General*, 131(2), 287.
- Konovalova, E., & Le Mens, G. (2017). Selective information sampling and the in-group heterogeneity effect. *Proceedings of the 39th annual conference of the Cognitive Science Society*, 39.
- Le Mens, G., Kareev, Y., & Avrahami, J. (2016). The evaluative advantage of novel alternatives: An information-sampling account. *Psychological science*, 27(2), 161–168.
- Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological review*, 97(2), 185.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4), 297–323.
- Von Neumann, J., & Morgenstern, O. (2007). *Theory of games and economic behavior*. Princeton university press.
- 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.