Experiencing Risk: Higher-order Risk Attitudes in Description- and Experience-based Decisions

Christoph K. Becker[1], Eyal Ert[2], Stefan T. Trautmann*[1,3] & Gijs van de Kuilen[3]

July 18, 2019

Abstract

Risky decisions are often characterized by imprecision about consequences and their likelihoods that can be reduced by information collection, and by unavoidable background risk. This paper addresses both aspects by eliciting risk attitude, prudence and temperance in decisions from description and decisions from experience. The results reveal a new description-experience gap for prudence, and replicate the known gap for risky decisions. While widespread prudence has been observed in decisions form description, we find no evidence of prudent decision making from experience. In decisions from experience, people are strongly influenced by the sampled mean, while standard deviation and skewness play a smaller role.

Highlights

- Elicitation of risk attitude, prudence and temperance in decisions from description and from experience
- Results reveal a new description-experience gap for prudence
- Sampled mean highly predictive of choices in experience-based decisions
- Sampled skewness and to lesser extend standard deviation predictive of choices in experience-based decisions

Keywords: higher-order risk attitudes, prudence, temperance, description-experience gap, sampling

---

[1] University of Heidelberg, [2] Hebrew University, [3] Tilburg University

* Corresponding author: Alfred-Weber-Institute for Economics, University of Heidelberg, Bergheimer Str. 58, 69115 Heidelberg, Germany, Phone: +49 6221 54 2952, Fax: +49 6221 54 3592; email: trautmann@uni-hd.de.

We thank audiences at the Behavioral Insurance Workshop in Munich, the Experimental Finance in Copenhagen, and the HeiKaMaxY Workshop in Karlsruhe for helpful comments.

1 Introduction

Uncertainty is a common component of many decisions, and people often try to avoid or reduce known uncertainties. Economists typically quantify these tendencies with measures of risk aversion. Avoiding a risk is, however, not always possible. Many decisions involve unavoidable background risks, either due to the specific environment (e.g. inflation), or due to other peoples' decisions (e.g., while driving). In these cases, measures of risk aversion are often insufficient to explain observed decisions, both theoretically and empirically (Noussair et al., 2014; White, 2008). Analyses of decision making need to consider higher-order risk attitudes in these cases, in particular *prudence* and *temperance* (Eeckhoudt and Kimball, 1992). Broadly speaking, prudence is the tendency towards precautionary behavior, while temperance is the propensity to disaggregate risks. Following Eeckhoudt and Schlesinger (2006), these higher risk attitudes can be elicited in a direct way, using simple binary lottery choices. Studies employing this method found that especially prudence is a widespread attitude, and a good predictor of behavior in decisions with unavoidable background risk; results are more mixed for temperance (Ebert and Wiesen, 2014; Noussair et al., 2014; Trautmann and van de Kuilen, 2018).

The way decision makers treat uncertainty depends also on the information about possible outcomes and their probabilities. According to Knight (1921), decision makers need to differentiate between *a-priori* probabilities (e.g. the probability of tails in a fair coin toss) and *statistical* probabilities, which are derived empirically. Outside of a small set of situations like the mentioned coin toss or casino gambling, most decisions fall into the second category. While there are different ways to infer probabilities empirically, the most straightforward one is to rely on past experiences. Every time a decision maker observes the consequences of a decision, new information is acquired that can be used the next instance a similar decision situation occurs. It is well known that decision makers' degree of risk aversion depends on whether information about a prospect is readily available, or whether it needs to be collected through experience and sampling. The difference between the two information formats is called the *description-experience gap* (Erev et al., 2010; Hau et al., 2010; Hertwig and Erev, 2009; Wulff et al., 2018). For example, the so called fourfold pattern of risk attitudes[1] reverses if there is no a-priori information and sampling is required. That is, decision making processes differ substantially depending on whether risk is described, or experienced by the

[1] The fourfold pattern of risk attitudes describes the following behavior: risk aversion for gains with high probability and losses with low probability; risk seeking for gains with low probability and losses with high probability (Tversky and Kahneman, 1992).

decision maker. In a recent meta-analysis, Wulff et al. (2018) summarize the main results of the experience-sampling paradigm: i) decision makers rely on small sets of experience, ii) there are no strong recency effects, iii) sampling error cannot fully explain behavior, and iv) decision makers tend to make decisions more in line with expected value maximization (compared to the description paradigm).[2]

Most risky decisions outside the decision analyst's lab take place in a context close to the decision from experience paradigm and concern trade-offs in the presence of background risk. In the latter case, higher order risk attitudes are predicted to matter most from a decision theoretic point of view (Brunnermeier et al., 2007; Ebert et al., 2017; Eeckhoudt and Gollier, 2005; Eeckhoudt and Kimball, 1992). Moreover, higher order moments of risky distributions, notably their skewness, have important implications for risk behavior in financial markets (de Roon and Karehnke, 2018; Harvey and Siddique, 2000). Higher order risk attitudes relate closely to these higher order moments. However, financial market outcomes are naturally perceived by most investors in an experience-based way, reducing the ecological validity of decision from description assessments of higher order risk attitude for these markets. Assessing the prevalence of prudence and temperance in situations with experience and sampling is therefore warranted. However, despite the recent wave of studies assessing higher order risk attitudes in various setups and subject pools, to the best of our knowledge there is no such assessment within the decision from experience paradigm.

The current paper aims to provide evidence regarding the role of experience sampling for higher order risk attitudes. We combine the separate research streams on decisions from experience and on higher order risk preference (in the description paradigm). We replicate the documented description-experience gap for risk aversion and find a new gap for prudence: While prudence is strong in the decisions from description paradigm (as reported in many studies by now, (see e.g. Deck and Schlesinger, 2014; Ebert and Wiesen, 2014; Maier and Rüger, 2011; Noussair et al., 2014; Trautmann and van de Kuilen, 2018), we find no evidence for prudence in decisions from experience. We are also the first to assess the explanatory power of an extensive set of statistical moments for decisions from experience: Employing the actual distributions observed at the individual level to predict risky choices, we evaluate the influence of the mean, standard deviation, skewness, and kurtosis on the participants' decisions. Our results indicate that differences in the mean are most predictive for decisions,

[2] A recency effect describes a decision maker's reliance on his or her most recent memories or pieces of information to make a decision. In sampling experiments, this corresponds to a participant relying on the most recent outcomes s/he sampled.

with a more modest role played by skewness and standard deviation. We explain our finding in terms of the decision processes in decision from experience situations.

2 Experimental Setup

We study risk aversion, prudence and temperance in both the description and the experience paradigms. Our setup uses a within-person design, counterbalanced along the description-experience dimension. Moreover, the design includes two between-person conditions, distinguished by the exogeneity, respectively endogeneity, of the number of samples drawn in the decision from experience tasks. In what follows, we first present the general experimental setup, then give details on the risky choice tasks, and then describe the experimental procedures and the subject pool.

2.1 Design

Participants made a set of binary choices in two within-person parts: A description part, in which they made these decisions from description, and an experience part, in which they made decisions from experience, using the sampling paradigm. Each of the two parts consisted of six different decision problems. Participants had to choose between two risky prospects, neutrally framed as "left" and "right" (see Appendix A, figures A2 and A3 for screenshots) in each decision problem. These six decision problems consisted of two risk aversion tasks, two prudence tasks and two temperance tasks, which are described in detail in Section 2.2. The order of the two parts was counterbalanced: participants' in the *description-first* order played the description part followed by the experience part, while in the *experience-first* order the reversed order was used.

In the description part, participants saw a full description of the probabilities and payoffs of both prospects. Hence, in each decision task they had full information concerning the choice at hand. The experience part used the sampling paradigm as described by Hertwig and Erev (2009), where participants received no description of the prospects but instead sampled them during a *sampling stage* that preceded a *decision stage*. Specifically, in the sampling stage participants sampled the two payoff distributions by clicking on one of two buttons on their screen. Each click on one of these buttons produced a result of the underlying outcome distribution of the chosen prospect. Participants could sample different number of times from the two lotteries if they wished (see details below), in any order. After sampling, participants' proceeded to the decision stage in which they had to choose one of the lotteries.

At the end of the experiment, one of the 12 decision tasks was randomly selected and played out for real. For example, if a participant chose a risky lottery in the first risk aversion task and this task was determined to be payoff-relevant, the computer drew one of the outcomes from the underlying distribution. The outcome of this task constituted the payoff of the participant.

Our study was furthermore divided into two between-subject conditions: *exogenous sampling*, in which participants had a fixed number of twenty samples per task (distributed over the two prospects according to the participant's sampling preference), and *endogenous sampling*, in which participants decided themselves how many samples to draw. In the exogenous sampling condition, the sampling phase ended automatically after the twentieth sample and participants entered the decision stage. This stage was highlighted by an explicit warning that the decision in this round was potentially payoff-relevant. In the endogenous sampling condition, participants had to sample at least five times, before they could proceed to the decision stage. This was implemented to ensure that participants had at least a rough understanding of the prospects they were about to make a choice between. The minimum of five samples was explicitly announced to the participants. Otherwise, participants could sample up to fifty times and could leave the sampling stage at any time after the fifth sample drawn. The maximum of fifty samples was not explicitly mentioned in the instructions, in order to allow participants to sample freely without setting a reference point for the appropriate number of samples. However, once participants had only five samples left, they saw a warning on each subsequent screen, showing the number of samples they have left. Once the fiftieth sample was drawn, participants proceeded automatically to the decision stage, as in the exogenous sampling condition.[3] The experiment was programmed using z-Tree 3.6.7 (Fischbacher, 2007).

2.2 Tasks

Each condition included six different binary choice tasks: two tasks for the elicitation of risk aversion, two for the elicitation of prudence and two for the elicitation of temperance. As participants played both the description and experience part, each of the 12 prospects in the second part they played was slightly modified by adding 50 Cents to every outcome in these prospects, to obtain two similar but slightly different sets of lotteries. This transformation

[3] Out of 282 participants, the following actually reached the fiftieth sample in the experience part: 19 in risk aversion task 1, 38 in risk aversion task 2, 27 in prudence task 1, 24 in prudence task 2, 25 in temperance task 1, and 29 in temperance task 2 (see Section 2.2. for the tasks).

ensured that the distributions of both options within each task was not repetitive yet kept the same underlying moments. The participants in the description-first condition encountered Set 1 (unmodified lotteries) in the description part, and Set 2 (modified lotteries) in the experience-part. The participants in the experience-first condition encountered Set 1 (unmodified lotteries) in the experience part, and Set 2 (modified lotteries) in the description task. Table 1 presents a full description of all prospects.

TABLE 1: Decision tasks

Task	Set 1			Set 2		
	Risky, imprudent, or intemperate		*Safe, prudent, or temperate*	*Risky, imprudent, or intemperate*		*Safe, prudent, or temperate*
Risk aversion task 1 (high prob.)	0.9 – 15€ 0.1 – 0€	vs.	13.5€	0.9 – 15.5€ 0.1 – 0.5€	vs.	14€
Risk aversion task 2 (low prob.)	0.1 – 15€ 0.9 – 0 €	vs.	1.5€	0.1 – 15.5€ 0.9 – 0.5€	vs.	2€
Prudence task 1	0.5 – 10€ 0.375 – 8€ 0.125 – 0€	vs.	0.375 – 12€ 0.5 – 6€ 0.125 – 4€	0.5 – 10.5€ 0.375 – 8.5€ 0.125 – 0.5€	vs.	0.375 – 12.5€ 0.5 – 6.5€ 0.125 – 4.5€
Prudence task 2 (non ES)	0.5 – 9€ 0.4 – 7.5€ 0.1 – 0€	vs.	0.4 – 11.5€ 0.5 – 6€ 0.1 – 3€	0.5 – 9.5€ 0.4 – 8€ 0.1 – 0.5€	vs.	0.4 – 12€ 0.5 – 6.5€ 0.1 – 3.5€
Temperance task 1	0.125 – 15€ 0.75 – 9€ 0.125 – 3€	vs.	0.5 – 12€ 0.5 – 6€	0.125 – 15.5€ 0.75 – 9.5€ 0.125 – 3.5€	vs.	0.5 – 12.5€ 0.5 – 6.5€
Temperance task 2	0.125 – 20€ 0.75 – 10€ 0.125 – 0€	vs.	0.5 – 15€ 0.5 – 5€	0.125 – 20.5€ 0.75 – 10.5€ 0.125 – 0.5€	vs.	0.5 – 15.5€ 0.5 – 5.5€

Each of the risk aversion tasks consisted of a two-outcome risky lottery and a safe payoff. In risk aversion task 1, the high outcome of the risky prospect was associated with a high probability of 90%, while in risk aversion task 2 the high outcome was realized only with a low probability of 10%. The safe payoff was adjusted accordingly to be equal to the expected payoff of the respective lottery. Hence, the lottery and the safe payoff differed only in their riskiness and the variability of the outcomes. The higher-order tasks were constructed according to the framework of Eeckhoudt and Schlesinger (2006). In particular, for the prudence tasks, an additional mean-zero risk was added to either the low outcome (imprudent option) or the high outcome (prudent option) of a two-outcome prospect. Additionally, we increased the upside risk of the prudent option in prudence task 2, which gave the prudent option an overall higher expected value and skewness. For the temperance tasks, two zero-mean risks were either added to the same outcome (intemperate option) or evenly distributed over the two possible outcomes (temperate option) of a two-outcome prospect. Importantly, all higher-order lotteries were reduced using reduction of compound lotteries (as shown in Table 1), since risk apportionment in the setup of Eeckhoudt and Schlesinger (2006) is not possible in sampling paradigms. This implies losing the sequential nature inherent in risk apportionment (see Trautmann and van de Kuilen, 2018, for a discussion of the reduced versus the risk-apportionment forms). To ensure comparability of the description and the experience conditions, we therefore used the same reduced format in the description task as well, to allow for a clean comparison between the two conditions.

2.3 Laboratory Procedures and Participants

In the laboratory, participants were seated in individual cubicles equipped with computers, the general instructions, a receipt for their payment for later use and a pen. We also provided them with an additional sheet, explaining the prospect format used (see Appendix B). This sheet was included to make sure that participants understood how prospect distributions work. The sheet also contained a comprehension question. Participants were given three outcome distributions and were asked to sketch the appropriate distribution graph themselves. Their answer was checked by one of the experimenters. If their sketch was correct, they received the instructions for the first part of the experiment and could proceed. Otherwise, they received additional explanation and were asked to try again before proceeding.

After finishing the first part of the experiment, participants received the instructions for the second part and had to confirm again that they have read them before proceeding. Once the participants had finished the second part, they were informed which part and task was chosen for them to be payoff-relevant. They also saw the randomly determined outcome of their choice in this task. They answered a short demographic questionnaire (asking for age, gender and field of study) and received their payoff in private.

Both conditions were run at the AWI lab in Germany. Participants were recruited from a pool of volunteers, mostly consisting of Heidelberg University students, using Hroot (Bock et al., 2014). Participants received a show-up fee of €3 and faced substantial and steep incentives with lottery payoffs ranging from €0 to €20. In the exogenous sampling condition, a total of 182 participants took part, 92 in description-first and 90 in experience-first order. The average age of the participants was 24 years (SD = 4.6), 53% of them were female, and 35% had some background in economics. Participants in the exogenous sampling condition earned €11.54 on average.

The endogenous sampling condition included 282 participants, with 175 in the experience-first part and 107 in the description-first part.[4] The average age of participants was 23 years (SD = 4.2), with 54% female and 32% having a background in economics. On average, participants earned €10.71. We deliberately included a higher number of participants in the endogenous sampling condition in order to account for the higher variation in samples drawn in the endogenous sampling process.

3 Hypotheses

Numerous studies have shown that the fourfold pattern for risk preferences typically observed in decision from description is reversed for decision from experience (Hertwig et al., 2004). For example, the often-found risk seeking behavior for small probability gains reverses to risk averse behavior in a sampling paradigm. Focusing only on the gain domain in our experiment, we expect to find the same pattern in our risk aversion tasks.

Hypothesis 1 (Replication): For the high probability gain risky task, a higher proportion of participants chooses the risk averse option in the description part compared to the experience part. For the low probability gain risky task, a lower proportion of participants chooses the risk averse option in the description part compared to the experience part.

For decisions involving prudent lotteries we expect a similar gap. Given the skewness properties of the two lotteries in these decision problems, the prudence tasks are similar in structure to the low probability gains. On the other hand, the increased complexity of the prudence task lotteries and the larger number of outcomes may induce subjects to sample more extensively in the endogenous sampling condition. However, as the description-experience gap for risk has been shown to be persistent even when the number of samples was

[4] The discrepancy between the numbers of participants in both treatments was introduced when we conducted more sessions of the experience first treatment due to scheduling issues.

increased (Hau et al., 2008; Ungemach et al., 2009) or when participants received stronger incentives to sample (Hau et al., 2008), we expect the same to hold for the prudence tasks. Moreover, if decision makers rely on small and possibly recent samples (Wulff et al., 2018), the structure of the lotteries induces under sampling of the attractive right branch of the prudent lottery and of the unattractive left branch of the imprudent lottery. This effect would similarly lead to prudent lotteries appearing less attractive to decision makers compared to the description case. Prudence task 2 tests this prediction by making the potentially underweighted (in decision from experience) upside risk of the prudent option more attractive.

Hypothesis 2: In both prudence tasks, a higher proportion of participants chooses the prudent option in the description part compared to the experience part.

Evidence on temperance in decisions from description is more mixed (Trautmann and van de Kuilen, 2018). For our current interest in description-experience differences, we observe that the lotteries in the temperance decision tasks are symmetric. Asymmetry seems an important ingredient in the description-experience gap, suggesting that differences may be less pronounced for temperance. On the other hand, the fatter tails for intemperance allow decision makers to observe more extreme outcomes, both good and bad ones, in small samples. With bad outcomes framed as losses with respect to average samples and loss aversion, we predict that the temperate option becomes more attractive under experience. This would be true for both endogenous and exogenous sampling.

Hypothesis 3: In both temperance tasks, a lower proportion of participants chooses the temperate option in the description part compared to the experience part.

In addition, we are interested in potential differences in the sampling behavior of participants in the two different conditions. First, previous studies endogenizing the number of samples found that participants only drew small samples, around and mostly below 20 (Ert and Trautmann, 2014; Hau et al., 2010; Wulff et al., 2018). We expect to find similar numbers for the risk aversion tasks. Since the higher-order tasks are more complex in regard of the possible number of outcomes in both lotteries, we expect participants to take this into consideration when deciding to stop sampling and proceed to the decision round. If the focus on recent small samples is relevant, this would not affect preferences though, as discussed above.

Hypothesis 4: Participants draw more samples in the higher-order risk tasks compared to the risk aversion tasks.

4 Results

4.1 The description-experience gap

Table 2 shows the pooled proportions of participants choosing the risk averse (safe), prudent, or temperate option in the different tasks in the description and experience conditions. In both conditions, we clearly replicate the description-experience gap for binary risky choices: For the high probability gain versus safe choice, the proportion of risk averse choices is higher for description (91% and 93% in the exogenous and endogenous treatments, respectively) compared to experience (47% and 39%). Conversely, in the low probability gain versus safe choice, the number of risk averse decisions increases when switching from description to experience (from 35% and 47% to 72% and 75%, respectively). Hence, we find a similar reversal for risky choice as previous studies (Hertwig and Erev, 2009): risk seeking for low probability high outcome lotteries in description-based choice reverts to risk aversion in experience-based choice; and risk aversion for high probability modest outcome lotteries reverts to risk seeking.

TABLE 2: Proportion (in % of participants) of safe, prudent and temperate choices

	A: Exogenous Sampling			**B: Endogenous Sampling**			**ΔGap AB**
	Description[1]	Experience[1]	Difference test[2]	Description[1]	Experience[1]	Difference test[2]	Difference test[3]
Risk aversion task 1 (high prob.)	0.91***	0.47	<0.001	0.93***	0.39**	<0.001	0.074
Risk aversion task 2 (low prob.)	0.35***	0.72***	<0.001	0.47	0.75***	<0.001	0.122
Prudence task 1	0.64***	0.52	0.020	0.68***	0.46	<0.001	0.123
Prudence task 2 (non-ES)	0.82***	0.53	<0.001	0.82***	0.52	<0.001	0.798
Temperance task 1	0.49	0.41*	0.140	0.43**	0.37***	0.169	0.748
Temperance task 2	0.46	0.45	0.752	0.47	0.47	1.000	0.790

[1]Significance according to a binomial test of the proportion being the result of indifference

(H_0: choice proportion = 0.5, *** p<0.001; ** p<0.01; * p<0.05)

[2]p-value is from a two-sample proportions test

[3]p-value is from an unpaired t-test

The results of the prudence tasks reveal a clear difference as well: A high proportion of the participants in both conditions chose the prudent option (64%-82%) from description, which corroborates previous findings (e.g., Trautmann and van de Kuilen, 2018). Yet, when they made decisions from experience, this clear preference disappeared (46%-53%). There is a higher proportion of prudent choices in prudence task 2 compared to prudence task 1 in the description parts, as expected due to our manipulation of the upside risk branch of the prudent lottery. In line with hypothesis 2, the proportion of prudent choices decreased to around 50% in the experience-based condition for prudence task 2, suggesting indeed the (positive) skewness receives less weight in the decisions from experience.

For temperance, the proportion of participants choosing the temperate option in both the description and the experience part, in both the exogenous and the endogenous sampling condition, is close to 50%, with a tendency towards intemperance. Results are consistent with previous findings for description-based decisions in the literature (temperance being weaker than prudence in most studies). In Temperance task 1 and mainly in the endogenous sampling condition, we find a stronger tendency towards intemperance. Overall, the proportions of temperate choices in the description and the experience parts do not significantly differ from each other, however. We find no evidence for a description-experience gap for lotteries that differ only in the fatness of their tails. Asymmetry seems to be an important aspect of description-experience gaps.

Importantly, the proportions of participants' choosing a specific option and the resulting gaps are strikingly similar in both sampling conditions for all tasks, despite the endogenous size of the sample in the second condition, and the resulting large differences in sample across subjects (see Section 4.2). To test the gap across conditions, we define the gap at the individual level by the difference in choice between the description and the experience parts. The difference tests for ΔGap AB in Table 2 indicate no significant differences between endogenous and exogenous. Moreover, results are robust to using between-subjects comparisons on the basis of the first-part decisions of the two different counterbalanced orders instead of within-subjects analysis.

4.2 Statistical Moments and Decisions from Experience

To better understand what factors underlie the participants' choices, we conduct a panel-probit analysis for the experience parts of both conditions. The dependent variable is the probability of choosing the risk averse, prudent or temperate option in a given task. The key variables of interest are the differences (between the two lotteries in a choice problem) in the four central moments of the respective lottery *as observed by the participant* in her sample:

mean, standard deviation, skewness and kurtosis. More precisely, we calculate for each participant the four above mentioned statistical moments based on their overall individual sample for both lotteries in a given task. For the safe option in risk aversion tasks 1 and 2, the usual measures for skewness and kurtosis are not defined, making it impossible to calculate a difference. In these cases, the skewness and kurtosis are normalized to zero. This normalization captures that the risky options have either a positive or negative skewness or kurtosis versus the safe one, which might make them differently favorable for the decision makers. The regression analysis also includes different controls. All regressions control for age, gender, the order in which a participant played both parts, having an economics background and the specific task through task fixed effects (*Controls*). While these controls are exogenous with respect to the decision making process, sampling behavior is endogenous. We control for sampling features by including the number of samples drawn (if endogenous sampling), the number of samples drawn from the right lottery (hence the safe, prudent or temperate option), and whether the participant saw all outcomes of both lotteries (Fox and Hadar, 2006). We call this set of variables *Sampling Controls*. In the endogenous sampling regressions, we also run a specification that includes the interaction of the number of samples drawn and the differences in the statistical moments, to detect whether differences in sampling affect the correlation between the moment differences and choice behavior. Standard errors are clustered at the subject level.

Table 3 shows the estimation results of these models. First notice that the coefficient sizes and directions are similar in both conditions. There is a strong association of choices with the subjectively sampled mean differences, both in terms of the regressors in the probit model, as well as the marginal effects. The association with the difference in standard deviation is negative and significant, except in model III for the exogenous sampling condition, where we find no significant main effect for standard deviation. The marginal effect sizes for the difference in standard deviation are substantially smaller than those for the difference in the sampled means ($p<0.01$ in all models). The association of choices with difference in sampled skewness is positive and significant in all models of both conditions. It has smaller marginal effects compared to the sampled mean difference in most conditions ($p<0.01$, $p=0.51$ for Model III), but larger effects than standard deviation difference in the endogenous sampling condition ($p<0.01$; no difference in the exogenous condition). Skewness seeking thus matters for experience-based decision , and apparently more so than the standard variation. The regressors for the difference in kurtosis are insignificant and close to zero. Considering the interaction effects in model III, we find significant effects for standard deviation and the mean: a larger number of samples amplifies the main effects. However, all interactions effects exhibit small effect sizes. This indicates that the observed differences are not driven by the interaction of larger samples and higher moment differences.

TABLE 3: Explaining Choices in Decisions from Experience

	Exogenous		**Endogenous**		
	I	II	I	II	III
Sampled distribution moments					
ΔMean	0.498**	0.473**	0.420**	0.389**	0.179**
	(0.079) [0.144]	(0.079) [0.135]	(0.043) [0.117]	(0.043) [0.101]	(0.063) [0.127]
ΔSD	-0.096**	-0.157**	-0.068**	-0.078**	0.001
	(0.030) [-0.028]	(0.035) [-0.045]	(0.020) [-0.019]	(0.026) [-0.020]	(0.037) [-0.027]
ΔSkewness	0.165**	0.152*	0.247**	0.221**	0.239**
	(0.037) [0.048]	(0.036) [0.043]	(0.033) [0.069]	(0.034) [0.057]	(0.067) [0.065]
ΔKurtosis	-0.004	-0.021	0.008	-0.012	-0.011
	(0.015) [-0.001]	(0.016) [-0.006]	(0.010) [0.002]	(0.011) [-0.003]	(0.033) [-0.003]
Interaction between number of samples and moment differences					
ΔMean^2					0.013**
					(0.003)
ΔSD^2					-0.004**
					(0.001)
$\Delta\text{Skewness}^2$					0.001
					(0.002)
$\Delta\text{Kurtosis}^2$					-0.000
					(0.001)
Sampling Controls[1]	No	Yes	No	Yes	Yes
Controls[2]	Yes	Yes	Yes	Yes	Yes
N	192	192	282	282	282

Notes: Probit regressions, clustered standard errors in parentheses, marginal effects in brackets (signs of the marginal effects are corroborated by unreported OLS regression analysis). Differences in mean, SD, skewness and kurtosis are defined as the subjectively sampled difference in the respective statistical moment between the right and the left lottery of a task. 1: Sampling controls include a dummy for whether or not the participant saw all possible outcomes, the number of samples from the right lottery and, for endogenous sampling, the total number of samples drawn. 2: Controls include the constant, age, gender, the order in which the parts were played, being an economist and a dummy for each task. *, **,*** indicate significance at the .05,.01 and .001 significance level.

Previous studies also explored the concept of recency as a potential contributor to the emergence of a description-experience gap. A participant exhibiting recency would rely more on the more recent samples to make a decision. This is usually motivated by memory decay while sampling information or by assuming that participants discount earlier obtained information. Results on recency have in general been mixed (Hau et al., 2008; Hertwig et al., 2004; Rakow et al., 2008; Ungemach et al., 2009; Wulff et al., 2018). We use three definitions for recency, as employed by Wulff et al. (2018), to study whether participants exhibited recency. According to these three definitions, the samples of the individual participants are split into a primacy and a recency sets. This is done by either i) splitting the complete individual sample into two halves (across), ii) splitting the sample in half for each of the two sampled lotteries (within), or iii) splitting the sample in two halves at the second switching point (mirror image), i.e. the first time an individual returns to a previously sampled lottery. We then calculate the sampled mean for each prospect in both the primacy and the recency sets. As discussed above, the mean is highly predictive of choices in decisions from experience. We then compare whether the sampled mean in the primacy or recency set is the better predictor of the observed choices by counting the number of correct predictions (see Appendix E for the corresponding figures). The results from this analysis indicate that both the recency and the primacy sets performed equally well in predicting actual choices when using the across and mirror image split definitions, in both the exogenous and the endogenous conditions. The within split definition, on the other hand, produced consistent recency effects in both conditions. Thus, the current data suggest only modest evidence for recency.

4.3 Sampling Behavior

Participants in the endogenous sampling condition chose themselves when to terminate sampling. Figure 1 shows the distribution of the samples drawn by the participants for each task, the solid lines mark the respective median amounts of samples. For the risk tasks, the median number of samples were 15 and 19, respectively. Participants drew on average slightly larger samples in the higher-order tasks, with medians ranging from 23-25 (all significant larger than the medians of the risk tasks at the 1% level, Wilcoxon matched-pair sign test) That is, participants did not come close to exhausting the full number of possible samples. While the higher order risk tasks exhibit a rather even distribution of samples, the distribution the risk aversion tasks is bimodal with peaks at the upper and the lower end. This is most likely caused by the comparison of a sure payoff with a heavily tailed lottery in the risk aversion tasks.

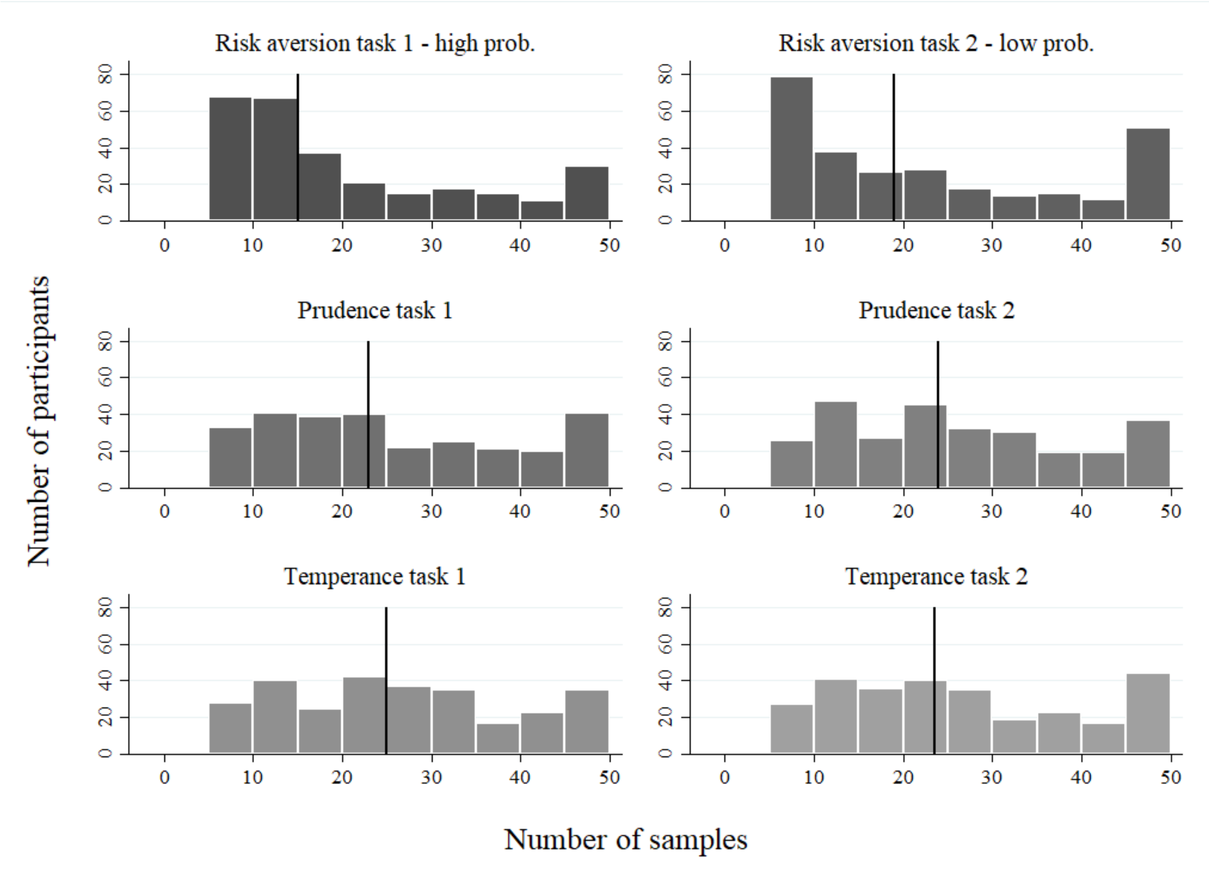


FIGURE 1: Distribution of samples drawn (black lines indicate the median)

Table 4 shows the description-experience gap split along the median number of samples drawn in the endogenous sampling condition. For both groups the proportion of risk averse, prudent or temperate decisions are close to each other in the description part, but vary in the experience part. For risk aversion, the gap is significantly larger for below median samplers. The effect points in the same direction for prudence, but insignificantly so. For temperance, the effect points in different direction in the two tasks (see column ΔGap AB in Table 4).

TABLE 4: Difference between below and above median samplers

	A: Below median samplers			**B: Above median samplers**			**ΔGap AB**
	Description[1]	Experience[1]	Difference test[2]	Description[1]	Experience[1]	Difference test[2]	Difference test[3]
Risk aversio task 1 (high prob.)	0.93***	0.21***	<0.001	0.92***	0.57	<0.001	<0.001
Risk aversion task 2 (low prob.)	0.49	0.86***	<0.001	0.45	0.65***	0.001	0.016
Prudence task 1	0.67***	0.43	<0.001	0.71***	0.49	<0.001	0.733
Prudence task 2 (non-ES)	0.82***	0.51	<0.001	0.81***	0.55	<0.001	0.548
Temperance task 1	0.44	0.41*	0.626	0.42	0.32***	0.100	0.392
Temperance task 2	0.54	0.44	0.109	0.41*	0.51	0.090	0.011

[1]Significance according to a binomial test of the proportion being the result of indifference

(H_0: choice proportion = 0.5, *** p<0.001; ** p<0.01; * p<0.05)

[2]p-value is from a two-sample proportions test

[3]p-value is from an unpaired t-test

Potentially, the observed gap may depend on whether participants saw all possible outcomes in each given task (Fox and Hadar, 2006). While a larger sample in general increases the likelihood of a participant observing all possible outcomes, there is still no guarantee that all outcomes are observed. Tables 5 and 6 therefore split the observations into two subsamples: one in which the participants did not see all possible outcomes of a task, and one in which they did. As the data from the exogenous condition can be split in the same way, these observations are included here. Table 5 shows the results for the exogenous sampling condition, while Table 6 shows the results for the endogenous sampling condition. The number of observations in each sample show that in many cases only about half of all participants did actually see all possible outcomes.

For the exogenous sampling condition in Table 5, no systematic effects emerge. The endogenous sampling condition in Table 6 exhibits a similar (but more pronounced) pattern as for splitting observations along the median of drawn samples presented in Table 4. This is not surprising as a larger sample increases the likelihood of observing all possible results (correlation ranging from 0.54 to 0.63 in the different tasks, Pearson's rho). The description-experience gap is substantially less pronounced for participants who did see all possible outcomes compared to those who did not, for both risk and the prudence tasks (see column ΔGap AB). Hence, sampling error at least partly drives the result in our data (Fox and Hadar, 2006).

TABLE 5: Exogenous - Difference between participants who saw all outcomes and those who did not

	A: Did not see all outcomes				**B: Saw all outcomes**				**ΔGap AB**
	N	Description[1]	Experience[1]	Difference test[2]	N	Description[1]	Experience[1]	Difference test[2]	Difference test[3]
Risk aversion task 1 (high prob.)	54	0.91***	0.47	<0.001	128	0.90***	0.46	<0.001	0.974
Risk aversion task 2 (low prob.)	94	0.42	0.69***	<0.001	88	0.28***	0.75***	<0.001	0.037
Prudence task 1	84	0.68**	0.55	0.071	98	0.60	0.48	0.133	0.870
Prudence task 2 (non-ES)	97	0.83***	0.55	<0.001	85	0.81***	0.51	<0.001	0.846
Temperance task 1	85	0.50	0.43*	0.353	97	0.48	0.40	0.250	0.919
Temperance task 2	84	0.49	0.39*	0.140	98	0.43	0.51	0.293	0.057

[1]Significance according to a binomial test of the proportion being the result of indifference

(H_0: choice proportion = 0.5, *** p<0.001; ** p<0.01; * p<0.05)

[2]p-value is from a two-sample proportions test

[3]p-value is from an unpaired t-test

The findings in tables 5 and 6 also explain why there are no differences in the description-experience gaps between the exogenous and endogenous sampling conditions: Participants with a very small sample in the experience task (which likely did not see all outcomes) appear to move further away from their choice in description on average. Conversely, those with a larger and more complete sample seem to be much closer to their choice in description. The overall choice pattern then remains close to that in the exogenous sampling condition. This explanation is supported by the stochastic properties of the prospects used. A simulation of hypothetical players sampling from the lotteries (see Appendix D) shows that a very small sample would lead to underestimation of the expected value of the safe option in risk aversion task 1, the risky option in risk aversion task 2, and the prudent option in prudence task 1 and 2. Given the large role of expected value for behavior in decision from experience (see Table 3), this pattern matches the description-experience gaps observed for risk and prudence.

TABLE 6: Endogenous - Difference between participants who saw all outcomes and those who did not

	A: Did not see all outcomes				**B: Saw all outcomes**				**ΔGap AB**
	N	Description[1]	Experience[1]	Difference test[2]	*N*	Description[1]	Experience[1]	Difference test[2]	Difference test[3]
Risk aversion task 1 (high prob.)	143	0.94***	0.13***	<0.001	139	0.91***	0.67***	<0.001	<0.001
Risk aversion task 2 (low prob.)	138	0.46	0.94***	<0.001	144	0.48	0.56	0.157	<0.001
Prudence task 1	133	0.67***	0.34***	<0.001	149	0.70***	0.58	0.030	0.008
Prudence task 2 (non-ES)	135	0.83***	0.36**	<0.001	147	0.82***	0.67***	0.003	<0.001
Temperance task 1	118	0.42	0.42*	1	164	0.43	0.34***	0.069	0.219
Temperance task 2	143	0.50	0.50	1	129	0.43	0.43	1	1

[1]Significance according to a binomial test of the proportion being the result of indifference

(H_0: choice proportion = 0.5, *** p<0.001; ** p<0.01; * p<0.05)

[2]p-value is from a two-sample proportions test

[3]p-value is from an unpaired t-test

5 Discussion & Conclusion

Research on higher-order risk attitudes has exclusively focused on decisions from description, despite the fact that such descriptions are rarely made in real life situations. Most situations involving risk taking are associated with at best partial knowledge about the underlying distributions, which are then experienced by observing outcomes over time. This study aims to test whether higher-order risk attitudes generalize to such situations of decisions from experience. In contrast to the widely demonstrated prevalence of prudence in decision from description, we find no significant prudence (nor imprudence) in decisions from experience.

That is, we find a substantial description-experience gap for prudence. We find no evidence for temperance, and no description-experience gap for temperance.

Our results also replicate the description-experience gap for risk attitudes found in previous studies. Hence, we assume that our experimental design is also well suited for studying the effect of decisions from experience on the higher-order risk attitudes. Skewed risks and higher-order risk attitudes have been shown to play an important role in economic and financial decisions (Kraus and Litzenberger, 1976; Noussair et al., 2014; Ebert and Hilpert, 2016; Drerup and Wibral, 2017). Especially prudence, as a measure for precaution in saving (Leland, 1968), bargaining (White, 2008), or auction bidding (Esö and White, 2004) has received close attention in economic research. The current results indicate that researchers need to critically evaluate the environment in which decisions take place. When information is not readily available and has to be acquired by experience, established behavioral patterns might vanish. Precautionary behavior and skewness seeking, as measured by prudence, seem less pronounced in these cases.

We use an extensive set of statistical moments to explain the participants' experienced-based choices. The results indicate that the subjectively sampled mean payoff is the most relevant predictor of choices between lotteries in the decisions from experience paradigm. Participants seem to have consistently chosen the lottery for which they observed the higher mean. While skewness seeking and aversion to standard deviation also significantly correlate with choices, their marginal effects are substantially smaller than for the effect of the mean; but skewness is more relevant than standard deviation for endogenous sampling. Moments like standard deviation, skewness and kurtosis, which are all measures for the variability of a distribution, should be intuitively easier to experience by sampling. Hence, we would expect a stronger effect. Still, our results are consistent with recent studies that also find a strong predictive effect of sampled means (Wulff et al., 2018). Hau et al. (2008) used several models and heuristics to explain results from experience-sampling experiments. The three best performing models were the maximax heuristic (choosing the lottery with the highest possible outcome), prospect theory with parameters fitted to past experience-sampling studies, and the natural mean heuristic (choosing the option with the higher mean). Erev et al. (2010) ran a choice-prediction competition to explain experimental data on experience-sampling choices. The natural mean heuristic once more performed remarkably well. Still, while our participants appear to be strongly influenced by the mean, we find that standard deviation and especially skewness also matter. However, we do not know what beliefs participants formed about the sampled distributions. Besides the subjective probabilities of the individual outcomes, participants might have formed beliefs about the existence of further, still unobserved outcomes.

Similar to other studies (Ert and Trautmann, 2014; Golan and Ert, 2015; Hau et al., 2010, 2008; Hertwig et al., 2004; Wulff et al., 2018), we find that participants rely on small samples. Participants drew slightly more samples when deciding between two non-degenerate lotteries, compared to deciding between a lottery and a safe option. Hence, participants seem to react to more complex decisions by adjusting the number of samples drawn, even though the overall sample is typically still small. Smaller samples in general carry a risk of missing information by not observing certain outcomes (Hertwig et al., 2006). This introduces a potential sampling error into participants' information sets that can distort their choices (Fox and Hadar, 2006). Indeed, we find that the gaps found in our data tend to be more pronounced for participants who sampled less in the endogenous sampling condition, and did not see all outcomes. That is, description-experience gaps seem to derive from a combination of sampling error and the decision to terminate sampling at a specific point in the endogenous sampling condition. Our result of no description-experience gap for the equally complex (in)temperance lotteries suggests that skewness is an important factor for the gap to occur.

These results leave us with further questions. Which criteria cause people to terminate sampling? Even in the endogenous sampling condition, there existed a minimum and a maximum number of samples. In decision situations outside the laboratory, there is typically little guidance as to how much information should be collected. In some practical settings, people receive information in the description format, but then experience outcomes by observing samples. This is typically the case with investment decisions, and it may lead to the ex-post unsatisfactory experience of outcomes, because the decision does not reflect the nature of the experience sampling after an investment made. One approach to account for this effect could be to shift experience to the stage were the decision is actually made, by letting people sample ex-ante from an already known distribution (Goldstein et al., 2008; Kaufmann et al., 2013). Such approaches seem especially promising in the presence of skewed risks as often found in investment and insurance decisions.

Literature

Bock, O., Baetge, I., Nicklisch, A., 2014. hroot: Hamburg registration and organization online tool. European Economic Review 71, 117–120.

Brunnermeier, M.K., Gollier, C., Parker, J.A., 2007. Optimal beliefs, asset prices, and the preference for skewed returns. American Economic Review 97, 159–165.

de Roon, F., Karehnke, P., 2018. Spanning tests for assets with option-like payoffs: The case of hedge funds. Available at SSRN 2667473.

Deck, C., Schlesinger, H., 2014. Consistency of higher order risk preferences. Econometrica 82, 1913–1943.

Drerup, T.H., Wibral, M., 2017. Skewness Expectations and Portfolio Choice. Working Paper, available at SSRN: https://ssrn.com/abstract=2810940 or http://dx.doi.org/10.2139/ssrn.2810940.

Ebert, S., Hilpert, C., 2016. Skewness Preference and the Popularity of Technical Analysis (SSRN Scholarly Paper No. ID 2354962). Social Science Research Network, Rochester, NY.

Ebert, S., Nocetti, D.C., Schlesinger, H., 2017. Greater mutual aggravation. Management Science 64, 2809–2811.

Ebert, S., Wiesen, D., 2014. Joint measurement of risk aversion, prudence, and temperance. Journal of Risk and Uncertainty 48, 231–252.

Eeckhoudt, L., Gollier, C., 2005. The impact of prudence on optimal prevention. Economic Theory 26, 989–994.

Eeckhoudt, L., Kimball, M., 1992. Background risk, prudence, and the demand for insurance, in: Contributions to Insurance Economics. Springer, pp. 239–254.

Eeckhoudt, L., Schlesinger, H., 2006. Putting risk in its proper place. The American Economic Review 96, 280–289.

Erev, I., Ert, E., Roth, A.E., Haruvy, E., Herzog, S.M., Hau, R., Hertwig, R., Stewart, T., West, R., Lebiere, C., 2010. A choice prediction competition: Choices from experience and from description. Journal of Behavioral Decision Making 23, 15–47.

Ert, E., Trautmann, S.T., 2014. Sampling experience reverses preferences for ambiguity. Journal of Risk and Uncertainty 49, 31–42.

Esö, P., White, L., 2004. Precautionary bidding in auctions. Econometrica 72, 77–92.

Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Experimental economics 10, 171–178.

Fox, C.R., Hadar, L., 2006. “ Decisions from experience”= sampling error+ prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004). Judgment and Decision Making 1, 159–161.

Golan, H., Ert, E., 2015. Pricing decisions from experience: The roles of information-acquisition and response modes. Cognition 136, 9–13.

Goldstein, D.G., Johnson, E.J., Sharpe, W.F., 2008. Choosing outcomes versus choosing products: Consumer-focused retirement investment advice. Journal of Consumer Research 35, 440–456.

Harvey, C.R., Siddique, A., 2000. Conditional skewness in asset pricing tests. The Journal of Finance 55, 1263–1295.

Hau, R., Pleskac, T.J., Hertwig, R., 2010. Decisions from experience and statistical probabilities: Why they trigger different choices than a priori probabilities. Journal of Behavioral Decision Making 23, 48–68.

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, 493–518.

Hertwig, R., Barron, G., Weber, E.U., Erev, I., 2006. The role of information sampling in risky choice. Information sampling and adaptive cognition 72–91.

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, 534–539.

Hertwig, R., Erev, I., 2009. The description–experience gap in risky choice. Trends in cognitive sciences 13, 517–523.

Kaufmann, C., Weber, M., Haisley, E., 2013. The role of experience sampling and graphical displays on one’s investment risk appetite. Management science 59, 323–340.

Knight, F.H., 1921. Risk, uncertainty and profit. New York: Hart, Schaffner and Marx.

Kraus, A., Litzenberger, R.H., 1976. Skewness preference and the valuation of risk assets. The Journal of finance 31, 1085–1100.

Leland, H.E., 1968. Saving and Uncertainty: The Precautionary Demand for Saving. The Quarterly Journal of Economics 82, 465–473.

Maier, J., Rüger, M., 2011. Experimental evidence on higher order risk preferences with real monetary losses. Chapman University (ESI), Working Paper 15–10.

Noussair, C.N., Trautmann, S.T., Van de Kuilen, G., 2014. Higher order risk attitudes, demographics, and financial decisions. The Review of Economic Studies 81, 325–355.

Rakow, T., Demes, K.A., Newell, B.R., 2008. Biased samples not mode of presentation: Re-examining the apparent underweighting of rare events in experience-based choice. Organizational Behavior and Human Decision Processes 106, 168–179.

Trautmann, S.T., van de Kuilen, G., 2018. Higher order risk attitudes: A review of experimental evidence. European Economic Review 103, 108–124.

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

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

White, L., 2008. Prudence in bargaining: The effect of uncertainty on bargaining outcomes. Games and Economic Behavior 62, 211–231.

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, 140.

Appendix

Appendix A: Presentation and Screen Shots

An example of how lotteries were depicted in the description task is depicted in Figure A1. Figure A2 and A3 show an example of the sampling procedure. Participants saw the two buttons "left" and "right". A click on them produced a randomly drawn result from the underlying distribution. Both pictures were taken from the endogenous sampling condition and hence also show the button to quit sampling (which appeared after sampling five times) and proceed to the final decision.

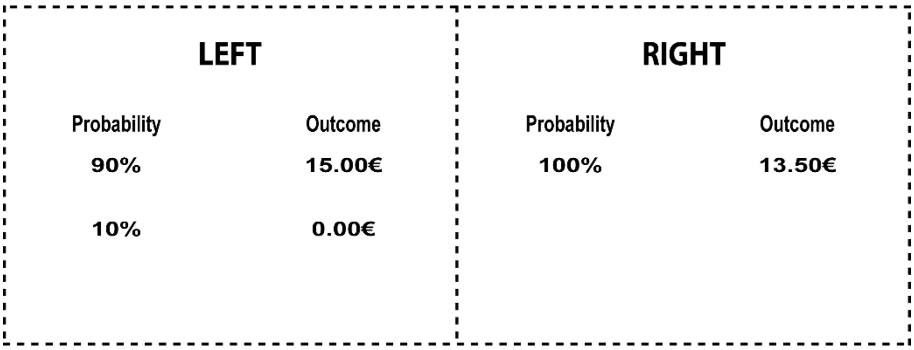


FIGURE A1: Example lottery – set 1, risk task 1

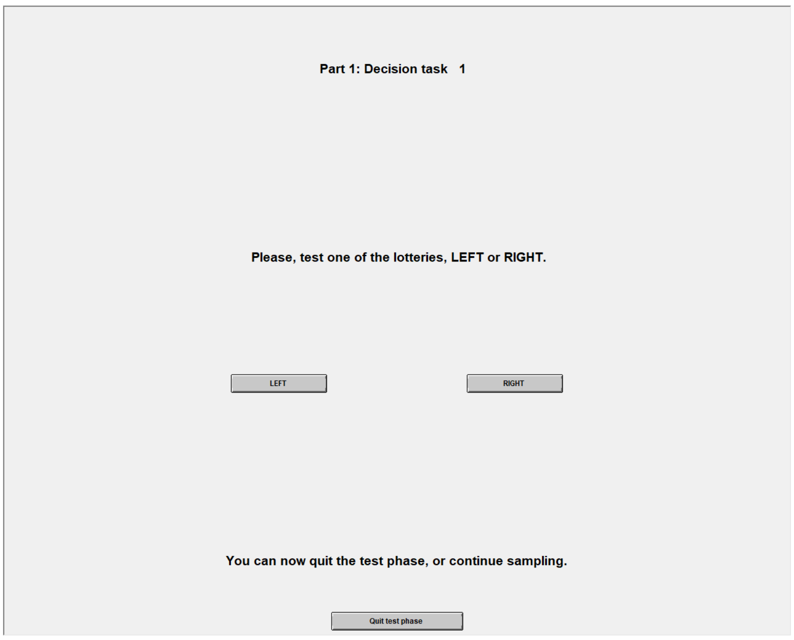


FIGURE A2: Sampling screen endogenous sampling: lottery buttons

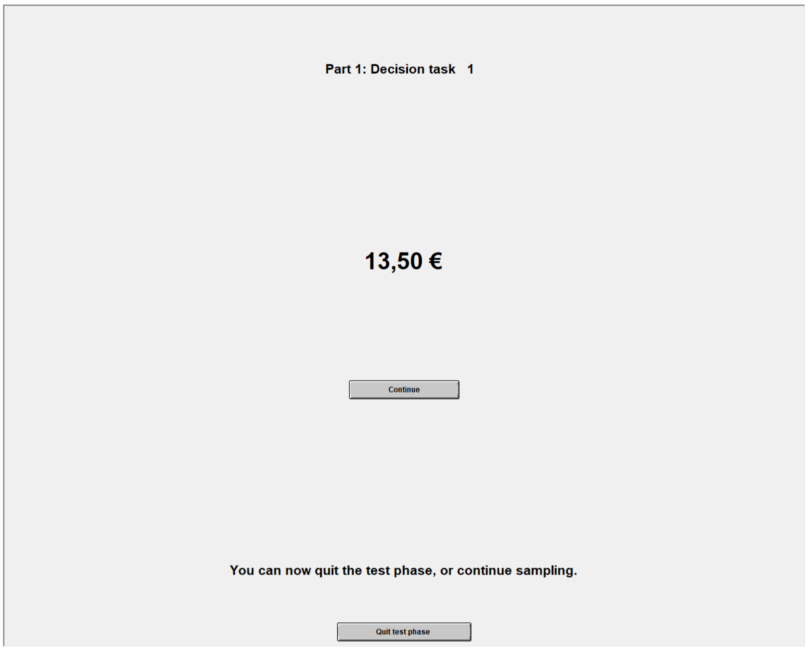


FIGURE A3: Sampling screen endogenous sampling: example outcome

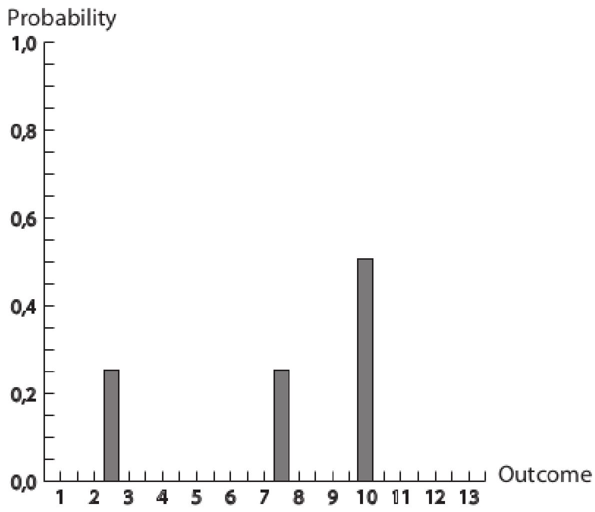
Appendix B: Comprehension Question

Example lottery

Please consider the following example risky lottery.
Such risky lotteries will play a central part in the experiment. Hence, we want to familiarize you with the concept.

Probability	Outcome
50%	10€
25%	7.5€
25%	2.5€

As a matter of illustration the diagram to the right depicts the possible outcomes and their probabilities of the lottery as a probability distribution.



Now consider the following lottery:

Probability	Outcome
25%	12.5€
25%	7.5€
50%	5€

Please draw the possible outcomes and their probabilities in the diagram to the right, as in the above example.

Once you finished this task, please raise your hand. One of the experimenters will come to your place to check your diagram. Please note that this task is only aimed to improve your understanding of the experiment. It is not relevant for your payoff. A rough draft will suffice.

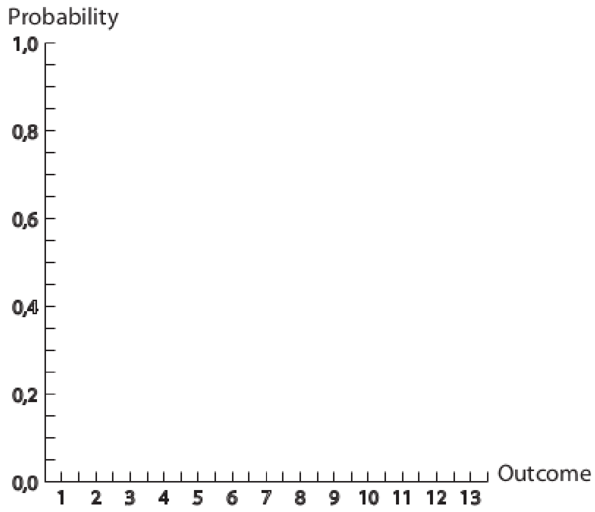


FIGURE A4: Example lottery with comprehension question, translated

Appendix C: Calculation of moments

For the calculation of the subjectively samples moments we used the inbuilt functions of Stata's *egen* command, namely the subcommands mean(), sd(), skew() and kurt(). These subcommands calculate the empirical measures for the mean, the standard deviations, the skewness and the kurtosis, respectively, according to the following formulas:

Mean:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Standard deviation:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Skewness (third moment):

$$m_3 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^3$$

Kurtosis (fourth moment):

$$m_4 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^4$$

Note: The standard deviation in both skewness and kurtosis is calculated with the following formula for the standard deviation: $s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$. This is because the calculation of skewness and kurtosis is defined over the calculation of distribution moments, with the rth moment of a distribution given as: $m_r = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s}\right)^r$.

Appendix D: Simulations

Figures A5-A10 show the results of simulating 100 hypothetical decision makers drawing samples from our different tasks. Each of these 100 decision makers was simulated to draw N=1, 2, 3, …, 50 samples from each of the two lotteries offered in each of the tasks. For every simulated set sampled in this way, the observed mean of each option was calculated. The sampled means of both options were then compared with each other. This procedure gives an overview of how a hypothetical sampler drawing equal-sized samples of size N from both options would have perceived the attractiveness of these options in terms of their mean. We ran 100 simulations (each with the aforementioned 100 samplers) and averaged over them. For simplicity we will only show graphs sampled from set 1 of our tasks (see Table 1).

In Figure A5, the safe option starts out with having a substantially lower sampled mean on average, compared to the risky option. This is to be expected: As the risk option offered a very high outcome (€15) with a large probability (90%), the risky option on average dominates the safe option with a payoff of €13.5 before the bad outcome occurs. Once a simulated sampler draws more samples, the likelihood of observing the low outcome (€0 with a 10% chance) increases, which reduces the likelihood of the risky option dominating the safe option. With more samples drawn, the likelihood of one option dominating the other in terms of the mean varies around 50%, with spikes around this value depending on the number of samples due to combinatorics. For risk task 2 the reverse pattern holds (Figure A6). For small number of samples, these patterns are consistent with the behavior observed for decisions from experience.

For prudence we observe that, similar to the case of the right-skewed risky lottery, for small samples the prudent lottery performs worse than the imprudent lottery in terms of expected value (figures A7 and A8). This is true for both prudence decision problems. Again, the pattern is consistent with the behavior observed for decisions from experience. For temperance no such effect is observed. This is consistent with the description-experience gap being closely related to the skewness of the lottery options.

FIGURE A5: Simulation of risk task 1

FIGURE A6: Simulation of risk task 2

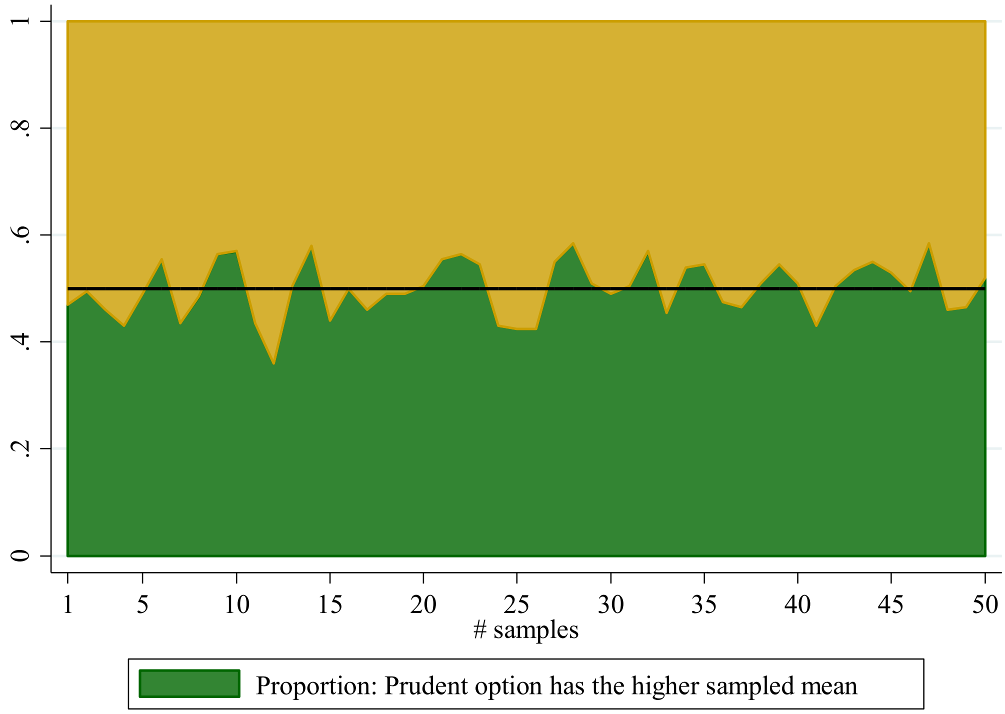


FIGURE A7: Simulation of prudence task 1

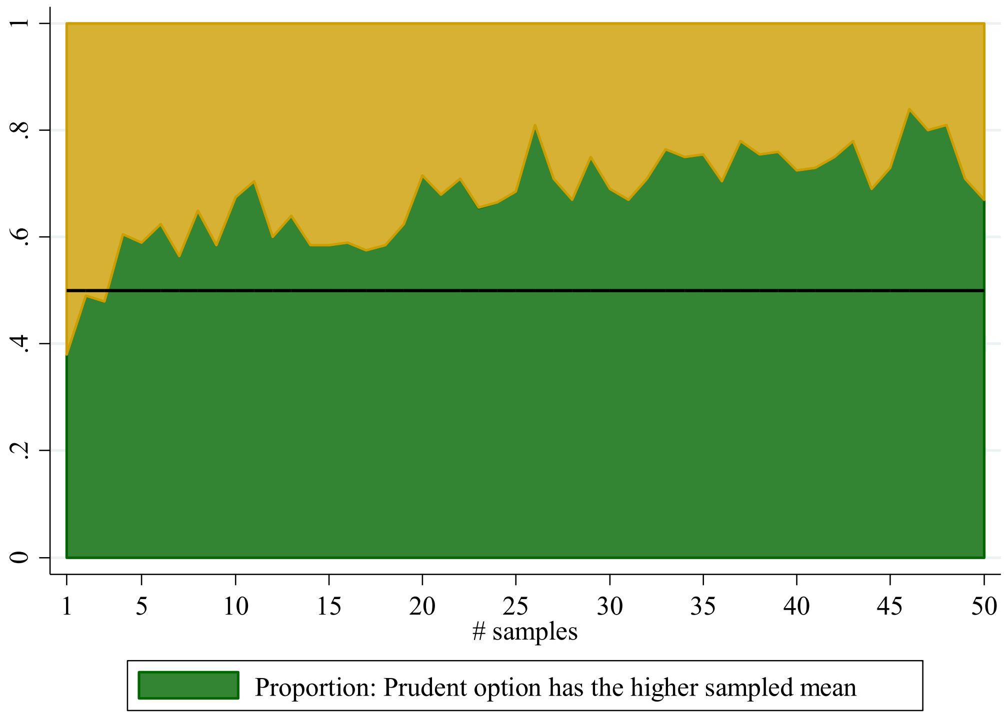


FIGURE A8: Simulation of prudence task 2

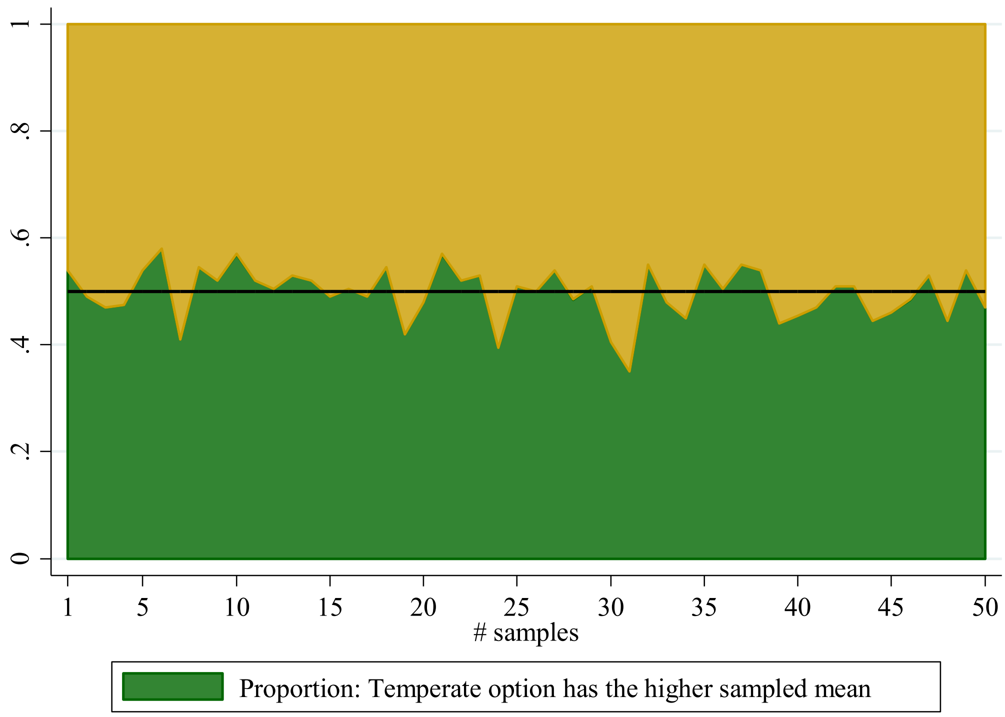


FIGURE A9: Simulation of temperance task 1

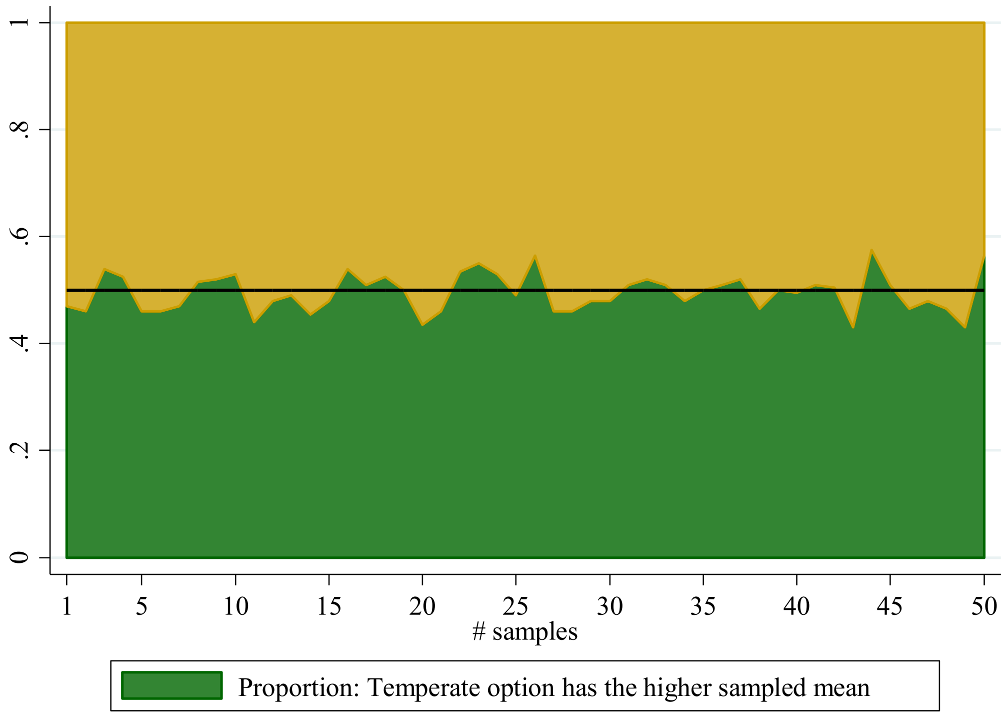


FIGURE A10: Simulation of temperance task 2

Appendix E: Recency and Primacy

Wulff et al. (2018) use three definitions for recency: For recency **across** lotteries, the sample is simply split in two halves (e.g. in the exogenous sampling condition, the first ten samples would constitute the primacy, the second 10 samples the recency set). Recency **within** lotteries is defined similarly, but the split is implemented within a given lottery (e.g. the first half of the samples of lottery Left in risk aversion task 1 are the primacy set, the second half of the samples from lottery Left form the recency set). Finally, the **mirror image** approach splits the samples at the point where the participant for the first time returns to sampling from the first lottery she sampled from. In other words, the split is implemented at the second switching point within the sample of each participant.

The figures below show a graphical representation of the degree of recency vs. primacy in our data. Figure A11 contains the results for the exogenous condition, Figure A12 the data for the endogenous condition. Each colored point in the graph for some decision task corresponds to one of the three definitions for splitting the data in a recency and a primacy set. These values are calculated as in Wulff et al. (2018) by subtracting the percentage of choices explained by the recency set from the percentage of choices explained by the primacy set. If both explained the observed choices to an equal degree, this difference should be zero.

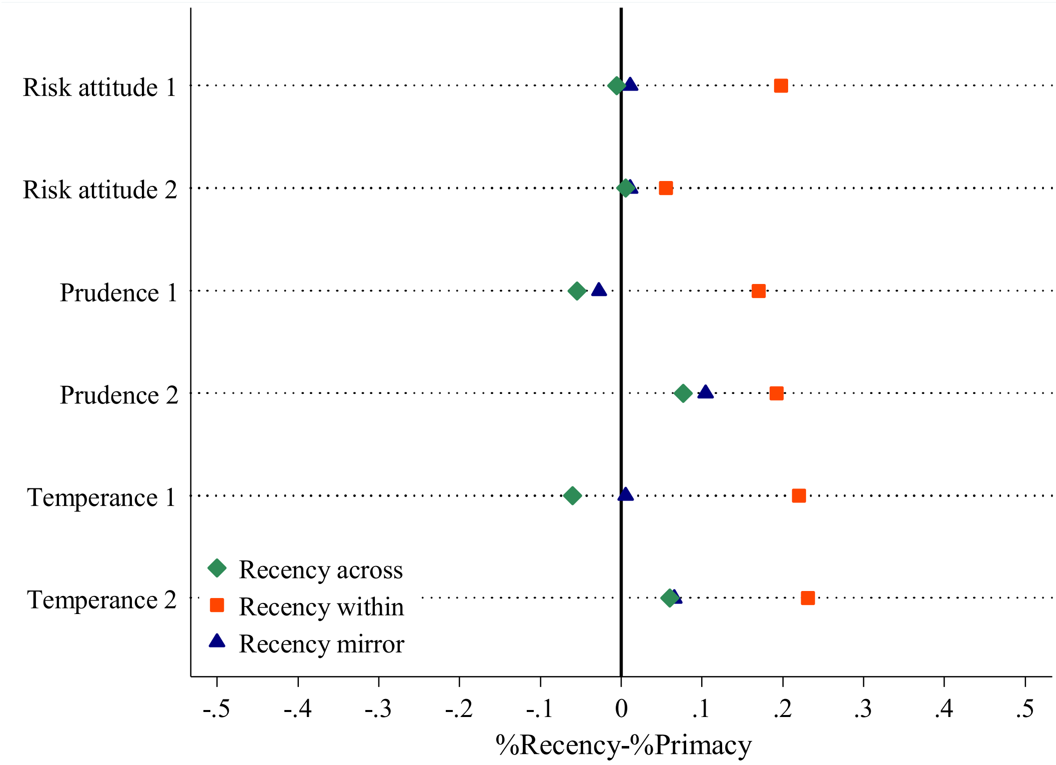


FIGURE A11: Recency in the exogenous sampling condition

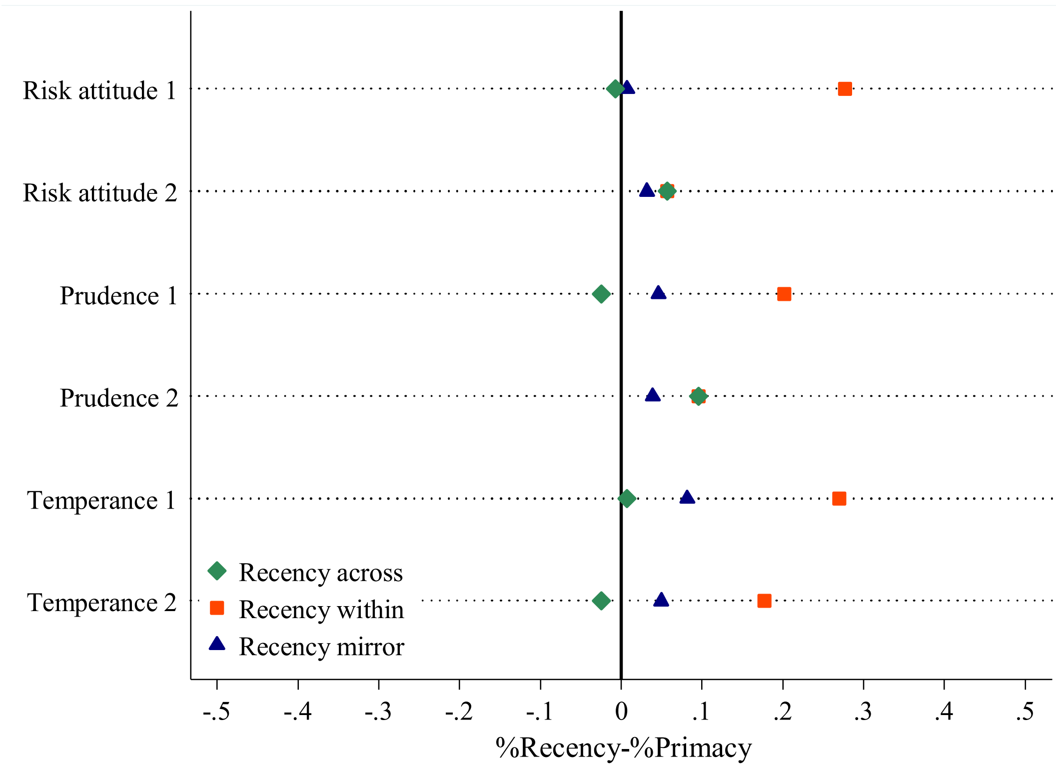


FIGURE A12: Recency in the endogenous sampling condition

Appendix F: Instructions

Appendix F1: General instructions

General Instructions

Dear participants,

Welcome and thank you for participating in our todays' experiment. Please, make sure that during the experiment, you:

- Do not talk with other participants.
- Switch off your mobile phone.
- Do not make any notes.

Should you have any questions during the experiment, please raise your hand. One of the experimenters will then come to your seat, to answer your question.

This experiment consists of two parts. In each part, you will make a decision in 6 different situations, henceforth called decision situations. At the end of the experiment, one of the two parts will be randomly selected. From this part, one decision situation will be randomly selected in turn and the outcome of your decision in this situation will be paid out to you. In addition, you will receive a safe payment of €3. Your total payoff at the end of the experiment thus amounts to:

Total payoff = Payoff of the randomly selected task of the randomly selected part + €3.

Before we start with the experiment, please examine the information sheet on uncertain lotteries and answer the comprehension question. Once you have done this, please raise your hand and one of the experimenters will subsequently check your answer. Afterwards, you will receive the instructions for Part 1 and can proceed with the experiment. Once you finished Part 1, please raise your hand again, in order to receive the instructions for Part 2. You will proceed through this experiment at your own speed, independent of the other participants.

The overall procedure of the experiment is depicted briefly on the back of these instructions.

Procedure

In each round, you will have to decide between LEFT and RIGHT. One randomly chosen decision situation from the two parts will determine your payoff at the end of the experiment.

General instructions + comprehension question

Raise your hand, to get the comprehension question checked and to get the instructions for Part 1.

Part 1

Raise your hand, to receive the instructions for Part 2.

Part 2

Random draws to determine the payoff relevant part and task.

Results screen

Fill in the receipt

Questionnaire

If you have questions, please raise your hand.

Appendix F2: Instructions exogenous sampling

Instructions Part 1

Part 1 of the experiment consist of 6 decision situations, in which you have to decide between the lotteries LEFT and RIGHT. Both lotteries will result in a payoff that respectively depends on a probability distribution. You can see an example of this below. You will not encounter this example as a decision situation in the experiment; it serves for illustrative purposes only.

Example decision situation

LEFT		RIGHT	
Probability	Payoff	Probability	Payoff
50%	**€10.00**	**25%**	**€12.50**
25%	**€7,50**	**25%**	**€7.50**
25%	**€2.50**	**50%**	**€5.00**

Please choose between LEFT and RIGHT

Would you choose LEFT in this example, you would receive €10 with a probability of 50%, €7.50 with a probability of 25%, and €2.50 with a probability of 25%. Would you choose RIGHT, you would receive €12.5 with a probability of 25%, €7.50 with a probability of 25% and €5 with a probability of 50%.

One of these 6 decision situations can be relevant for your payoff, as mentioned before. After you made all 6 decisions, the computer will randomly draw one of the decision situations. All decision situations have the same probability to be picked. The computer will then determine your payoff given your chosen option (LEFT or RIGHT) and the corresponding probability distribution. This result is your actual payoff for Part 1 and will be saved by the computer. If Part 1 is randomly selected at the end of the experiment to be payoff relevant, this result will be paid out to you.

Please raise your hand, if you have further questions for the experimenters. If you have no questions, you may begin with Part 1 at the computer.

Instructions Part 2

Part 2 of the experiment consists of 6 decision situations concerning uncertain lotteries. Each decision situation consists of testing stage of 20 rounds (not payoff relevant) and a subsequent decision round (potentially payoff relevant). These aspects will now be described in further detail.

Lotteries: In each decision situation, you will make a decision between the lotteries LEFT and RIGHT, without having initial information concerning the respective probability distribution and the possible outcomes. You can collect this information in the testing stage.

Testing stage: The testing stage consists of 20 rounds in which you can test the lotteries LEFT and RIGHT to draw a sample from the underlying (unknown) probability distribution. Immediately after each testing round, you will see the outcome of your chosen lottery. The samples are independent from each other and will be drawn with replacement (i.e. always from the same underlying probability distribution). The outcomes of the testing stage are not payoff relevant. After 20 rounds, you will receive an announcement that the testing stage is over.

Decision round: The decision round in each decision situation follows directly the respective testing stage. In the decision round you will make a singular, potentially payoff relevant decision between the lotteries LEFT and RIGHT based on the information collected from both lotteries in the testing stage.

One of the 6 decision situations of Part 2 can be relevant for your payoff, as mentioned before. After you made all 6 decisions, the computer will randomly draw one of the decision situations. All decision situations have the same probability to be picked. The computer will then determine your payoff given your chosen option (LEFT or RIGHT) and the corresponding probability distribution. This result is your actual payoff for Part 2 and will be saved by the computer. If Part 2 is randomly selected at the end of the experiment to be payoff relevant, this result will be paid out to you.

Please raise your hand, if you have further questions for the experimenters. If you have no questions, you may begin with Part 2 at the computer.

Appendix F2: Instructions endogenous sampling

Instructions Part 1

Part 1 of the experiment consist of 6 decision situations, in which you have to decide between the lotteries LEFT and RIGHT. Both lotteries will result in a payoff that respectively depends on a probability distribution. You can see an example of this below. You will not encounter this example as a decision situation in the experiment; it serves for illustrative purposes only.

Example decision situation

LEFT		RIGHT	
Probability	Payoff	Probability	Payoff
50%	**€10.00**	**25%**	**€12.50**
25%	**€7,50**	**25%**	**€7.50**
25%	**€2.50**	**50%**	**€5.00**

Please choose between LEFT and RIGHT

Would you choose LEFT in this example, you would receive €10 with a probability of 50%, €7.50 with a probability of 25%, and €2.50 with a probability of 25%. Would you choose RIGHT, you would receive €12.5 with a probability of 25%, €7.50 with a probability of 25% and €5 with a probability of 50%.

One of these 6 decision situations can be relevant for your payoff, as mentioned before. After you made all 6 decisions, the computer will randomly draw one of the decision situations. All decision situations have the same probability to be picked. The computer will then determine your payoff given your chosen option (LEFT or RIGHT) and the corresponding probability distribution. This result is your actual payoff for Part 1 and will be saved by the computer. If Part 1 is randomly selected at the end of the experiment to be payoff relevant, this result will be paid out to you.

Please raise your hand, if you have further questions for the experimenters. If you have no questions, you may begin with Part 1 at the computer.

Instructions Part 2

Part 2 of the experiment consists of 6 decision situations concerning uncertain lotteries. Each decision situation consists of testing stage over multiple rounds (not payoff relevant) and a subsequent decision round (potentially payoff relevant). These aspects will now be described in further detail.

Lotteries: In each decision situation, you will make a decision between the lotteries LEFT and RIGHT, <u>without</u> having initial information concerning the respective probability distribution and the possible outcomes. You can collect this information in the testing stage

Testing stage: In the testing stage, you can test the lotteries LEFT and RIGHT to draw a sample from the underlying (unknown) probability distribution. Immediately after each testing round, you will see the outcome of your chosen lottery. The samples are independent from each other and will be drawn with replacement (i.e. always from the same underlying probability distribution). The outcomes of the testing stage are not payoff relevant. From the **fifth** testing round onwards, you gain the possibility to quit the testing phase and proceed to the decision round. Alternatively, you can further test the lotteries. The possibility to proceed to the decision round can be used any time after the fifth round.

Decision round: The decision round in each decision situation follows directly the respective testing stage, after you quit the testing stage. In the decision round you will make a singular, potentially payoff relevant decision between the lotteries LEFT and RIGHT based on the information collected from both lotteries in the testing stage.

One of the 6 decision situations of Part 2 can be relevant for your payoff, as mentioned before. After you made all 6 decisions, the computer will randomly draw one of the decision situations. All decision situations have the same probability to be picked. The computer will then determine your payoff given your chosen option (LEFT or RIGHT) and the corresponding probability distribution. This result is your actual payoff for Part 2 and will be saved by the computer. If Part 2 is randomly selected at the end of the experiment to be payoff relevant, this result will be paid out to you.

Please raise your hand, if you have further questions for the experimenters. If you have no questions, you may begin with Part 2 at the computer.