Probability, but not utility, influences repeated mental simulations of risky events

Yun-Xiao Li (yunxiao.li@warwick.ac.uk)[1], **Johanna K. Falbén (j.falben@uva.nl)**[1,2], **Lucas Castillo (lucas.castillo-marti@warwick.ac.uk)**[1], **Jake Spicer (jake.spicer@warwick.ac.uk)**[1], **Jian-Qiao Zhu (jz5204@princeton.edu)**[1,3], **Nick Chater (nick.chater@wbs.ac.uk)**[4], and **Adam N. Sanborn (A.N.Sanborn@warwick.ac.uk)**[1]

[1]Department of Psychology, University of Warwick, [2]Department of Psychology, University of Amsterdam, [3]Department of Computer Science, Princeton University, [4]Warwick Business School, University of Warwick

Abstract

There has been considerable interest exploring how the utility of an outcome impacts the probability with which it is mentally simulated. Earlier studies using varying methodologies have yielded divergent conclusions with different directions of the influence. To directly examine such mental process, we employed a random generation paradigm in which all the outcomes were either equally (i.e., followed a uniform distribution) or unequally (i.e., a binomial distribution) probable. While our results revealed individual differences in how the utility influenced responses, the overall findings suggested that it is the outcomes' probabilities, not their utilities, that guide this process. Notably, an initial utility-independent bias emerged, with individuals displaying a tendency to start with smaller values when all outcomes are equally likely. Our findings offer insights into the benefits of studying the mental sampling processes and provide empirical support for particular sampling models in this domain.

Keywords: random generation; sampling; probability judgment

Introduction

Imagine purchasing a lottery ticket—what results do you envision? Psychologists have delved into understanding how people perceive risky events, developing sequential sampling models such as Decision Field Theory (DFT; Busemeyer & Townsend, 1993; Busemeyer & Diederich, 2002; Roe, Busemeyer, & Townsend, 2001) and Utility-Weighted Sampling (UWS; Lieder, Griffiths, & Hsu, 2018). In the face of a risky event, such as an option with a fifty-fifty chance of winning either £3 or £5, these models assume individuals mentally simulate or "sample" potential outcomes from a probability distribution.

However, DFT and UWS make different assumptions about the interplay between the utilities of outcomes and the probabilities of them being sampled. DFT makes the simple assumption that the probability of sampling an outcome is the probability described to participants. To illustrate, in the aforementioned example, individuals would draw a roughly equivalent number of samples for both £3 and £5 outcomes. In contrast, UWS assumes "oversampling" of the extreme outcomes, i.e., individuals would draw a greater number of samples for the larger (£5) outcome.

Past investigations on whether utilities impact the perception of their probabilities have generally found that this is indeed the case. However, these past investigations do not agree about *how* outcomes affect probabilities. By characterising probability judgments, choices and recall as driven by a mental sampling process (Zhu, León-Villagrá, Chater, & Sanborn, 2022; Zhu, Sundh, Spicer, Chater, & Sanborn, 2023), these findings can be categorised into four types of "biases" considering the oversampling or undersampling of extreme gains and extreme losses, as illustrated in Figure 1.

Supporting the assumptions of UWS, some empirical studies in decisions from experience found that extreme gains and extreme losses were overestimated (Ludvig, Madan, & Spetch, 2014; Madan, Spetch, Machado, Mason, & Ludvig, 2021) and were more likely to be recalled (Madan, Ludvig, & Spetch, 2014). We describe this bias as *polarisation* (see Figure 1).

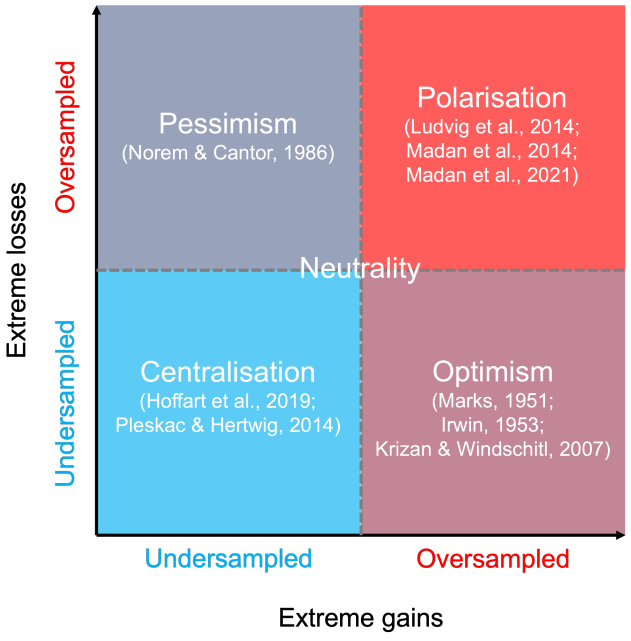

Figure 1: Potential biases regarding the sampling of extreme gains and extreme losses, with citations of studies supporting each bias.

Conversely, Pleskac and Hertwig (2014; see also Hoffart, Rieskamp, & Dutilh, 2019) found that participants estimated lower probabilities associated with better outcomes, suggesting a *centralisation* bias where larger rewards or losses were

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

perceived as less likely than smaller ones (also termed a "risk-reward heuristic").

Individuals were also found to exhibit variations in their levels of *optimism* or *pessimism* across different tasks. Studies have demonstrated that participants tended to perceive desirable events as more likely to occur and undesirable events as less likely—both perceptions often deviating from the descriptive probability (Irwin, 1953; Marks, 1951; Krizan & Windschitl, 2007). Interestingly, in a performance prediction task, Norem and Cantor (1986) revealed a group of "defensive pessimists" who reliably expected poorer performance that those identified as optimists.

Despite exploring the interaction between utilities and probability judgments using varied procedures, these studies did not directly investigate the sampling process itself, on which the probability judgments rely. And, importantly, past work has generally not considered the time course of biases: While past work has characterized these biases as constant, it is possible that just the first few mental simulations are biased by utilities. For example, mental sampling accounts of the anchoring effect assume that the anchor affects only the starting point of the mental simulation process, with later samples gradually becoming unbiased (Lieder, Griffiths, Huys, & Goodman, 2018). Notably, the individual level has received limited attention in these studies, and the presence of outliers could introduce bias into the results.

To examine the time course of biases in mental simulation, we adapted a paradigm that calls for repeated mental simulation: the random generation task (Baddeley, 1966). In this task participants were asked to repeatedly mentally simulate playing a lottery and utter the outcome out loud while following the beats of a metronome. Random generation was initially used to study working memory (Baddeley, 1998) and people's concept of randomness (Nickerson, 2002), and has recently been adapted to explore the mental sampling process itself, both for equally probable and unequally probable events (Castillo, León-Villagrá, Chater, & Sanborn, 2024; León-Villagrá, Castillo, Chater, & Sanborn, 2022).

In general, we tested for the hypotheses depicted in Figure 1 using the random generation paradigm with two probability distributions: a uniform distribution (i.e., drawing slips of papers from a bag; Experiment 1) and a binomial distribution (i.e., the number of heads after tossing ten coins simultaneously; Experiment 2).

Experiment 1

Participants

28 participants (8 women, 20 men; age: M = 21.9 years, SD = 3.5) took part in the study. Two additional participants were recruited but were excluded from the analysis since the length of their generated sequence was shorter than 80% of the requested length. Participants received a show-up fee of £3, plus a bonus of up to 20p depending on the number of points they received at the end of the experiment. The average payment was £3.10, and the experiment took about 40 minutes to complete.

Design and Materials

The experiment employed a within-subject design, incorporating three variants of a random generation task through three distinct scenarios: winning, losing, and control. All the participants began with the control scenario, and the order of the subsequent winning and losing scenarios was counterbalanced across subjects.

In each scenario, participants were presented with a set of 11 slips of paper of identical size reflecting the numbers 0 to 10. Scenarios differed in the value of these numbers: In the winning scenario, slips were labelled as "winning [x] points"; in the losing scenario, slips were labeled as "losing [x] points", where "x" represents each of the numbers 0 to 10; in the control scenario, slips were only labeled with their respective number. Participants were asked to imagine "drawing a slip out from a bag, saying the number or the outcome on it out loud, putting it back, shuffling, then repeating the process", following the instructions in Baddeley (1966). Notably, in the winning and losing scenarios, participants were asked to articulate the full phrase "winning" or "losing [x] points" during every response to ensure scenario retention.

To aid comparison between scenarios, each generated response was converted into its absolute value, meaning "5" in the control scenario, "winning 5 points" in the winning scenario, and "losing 5 points" in the losing scenario were all transcribed as 5 (and all termed "number" below).

Procedure

The experiment was conducted online using Microsoft Teams, which was also used to record participants' audio. The experiment began with an explanation of the random generation task, including a visual demonstration of the drawing process using physical slips by the experimenter.

The experiment contained three blocks, one for each of the three scenarios. Each block began by showing a photograph of the set of slips from that scenario. Participants were then asked to generate draws from that set as spoken responses for 4 minutes. To guide the pace of responses, participants were presented with a flashing dot in the middle of their screen, appearing 37 times per minute. Participants were asked to generate a number every time the dot appeared. They produced responses slightly faster than requested, producing an average of 156.77 numbers (SD = 27.82).

At the end of each block, participants received a questionnaire with two questions: first, a forced-choice question, whether all the numbers were equally likely; and second, the probabilities of each number as a percentage without requiring that they sum to 100. After the questionnaire, in the winning and losing scenarios, the experimenter took a slip from a bag in front of the participants. The outcome on the slip was either added or subtracted from the point total, and this total was then converted into a cash bonus, with each point equal-

ing one penny. [1] This approach aimed to encode outcomes as either losses or gains while keeping them independent from the random generation tasks.

Results

Distribution of Responses We analyzed participants' responses to the forced-choice question: The proportions of participants selecting "yes" were significantly higher than 50% in all three scenarios (Control: 22/28, $Z = 2.83, p = .002$; Winning: 23/28, $Z = 3.21, p < .001$; Losing: 19/28, $Z = 1.70, p = .044$) after a Holm-Bonferroni correction at the significance level of .05.

Figure 2A illustrates the distribution of (normalised) probability judgments and random generation proportions across numbers for each scenario. Comparisons with the uniform distribution via one-sample t-tests revealed tendencies to underestimate 0, under-generate 0 and 1 in the winning scenario, and over-generate 3 in the losing scenario. However, these statistical differences were explained by individual differences found in further analysis. Furthermore, we observed high similarity between the probability judgment and random generation distributions, quantified by the overlapping coefficient (Pastore & Calcagnì, 2019; Weitzman, 1970) measuring the percentage of these distributions covering the same area; overlapping coefficients were calculated individually for each subject, and the mean coefficient for each scenario is presented in Figure 2A.

Model Comparison of Random Generation Data To examine whether the scenarios influenced the means of the generated numbers, we conducted an extensive comparison of Bayesian linear mixed models.[2] Table 1 shows the set of models and their natural log of Bayes factors (lnBF). Because Model 0 was used as the reference model in each Bayes factor, the lnBF comparing any other two models can also be calculated by taking the difference of the lnBFs reported in the table. Further, because of the common reference model, the model with the highest lnBF provides the best predictions.

Model 9 performed the best among these models, implying individual differences, but no fixed effects, across the influence of the scenarios on the average of the numbers generated (as Model 9 outperformed Model 10; $BF = 160.77$). In addition, as Model 9 outperformed Model 1 ($BF = 5.84 \times 10^6$), the first generated number (i.e., the 'starting point') appears to deviate from the overall mean. However, such starting-point bias is utility-independent as Model 11 underperformed Model 10 ($BF = 9.28 \times 10^{-3}$).

Individual Differences in the Effect of Scenarios We next explored the individual differences in the effect of the scenario identified in the Bayesian mixed linear model comparison. We first tested each participant against the neutrality hypothesis: In the winning and losing scenarios, each person drew samples from the same distribution as they did in the control scenario. We thus calculated the mean values of each scenario for each participant, and then subtracted the mean value of the control scenario from the mean values of the winning and the losing scenarios. Figure 3A shows the relationship between the empirical winning-control mean difference against the empirical losing-control difference on the individual level.

In order to test the neutrality hypothesis, we conducted a Monte-Carlo (MC) simulation by drawing samples from the random generation response distribution in the control scenario (shown as the blue bar of the left panel of Figure 2A).[3] The MC simulation can provide a likelihood estimation of the neutrality hypothesis, and thus can also provide a 95% confidence region, which is shown as a red ellipse in Figure 3A. We found that 23 out of 28 participants' points (82%) fell in the ellipse, significantly more than 50% ($p < .001$, 95% CI=[63.11%, 93.94%]). Thus, more than half of the participants were not influenced by the scenarios. Of the remaining participants, there were some who showed optimism, centralisation, or polarisation, but none who showed pessimism.

Of the hypotheses of bias proposed in past work, only the UWS hypothesis (e.g., polarisation in Figure 1), has been specified well enough to quantitatively compare against neutrality in our data. In their original paper, Lieder, Griffiths, and Hsu (2018) defined the utility of an outcome o as

$$u(o) = \frac{o^k}{o_{max}^k - o_{min}^k} + \varepsilon, \tag{1}$$

where k was set to $k = 1$, o_{max} and o_{min} were the largest and the smallest outcome in the given choice, and $\varepsilon \sim N(0, \sigma)$ was neural noise (omitted in our analysis). We also tested a utility function with $k = 0.5$ as the range of 0.5 to 1 covers the range used in previous studies (e.g., Glöckner & Pachur, 2012).

To calculate the likelihood of UWS, we employed a similar MC method.[3] Therefore, the neutrality hypothesis and both UWS models have their respective likelihood estimations on the plane depicted by Figure 3A. Then, we found the best-fit hypothesis according to their likelihoods for each participant. The neutrality hypothesis outperformed both UWS models for 27 participants, and the UWS with the exponent of 0.5 best fit the remaining participant. These findings suggest that the utilities of the outcomes did not have an impact on the sampling process as predicted by UWS.

Fixed Effect of Starting-point Bias As indicated in the model comparison, participants exhibited a starting-point bias. To investigate the direction people were biased, we calculated the mean values of the first five numbers generated in each of the three scenarios (see Figure 4). Participants tended to begin with relatively smaller numbers but within a few responses converged on the theoretical mean of 5. Although the

[1] In the losing scenario, participants were endowed with 10 points before the gamble so they would not truly lose points.

[2] For more details about the estimation of the Bayesian linear mixed models and their lnBF, please refer to the Appendix 1 on OSF (https://osf.io/a6w3v/?view_only=7ee3219936b647c1a63c796296072cde).

[3] For more details about the MC simulation, please refer to the Appendix 2 on OSF (link in Footnote 2).

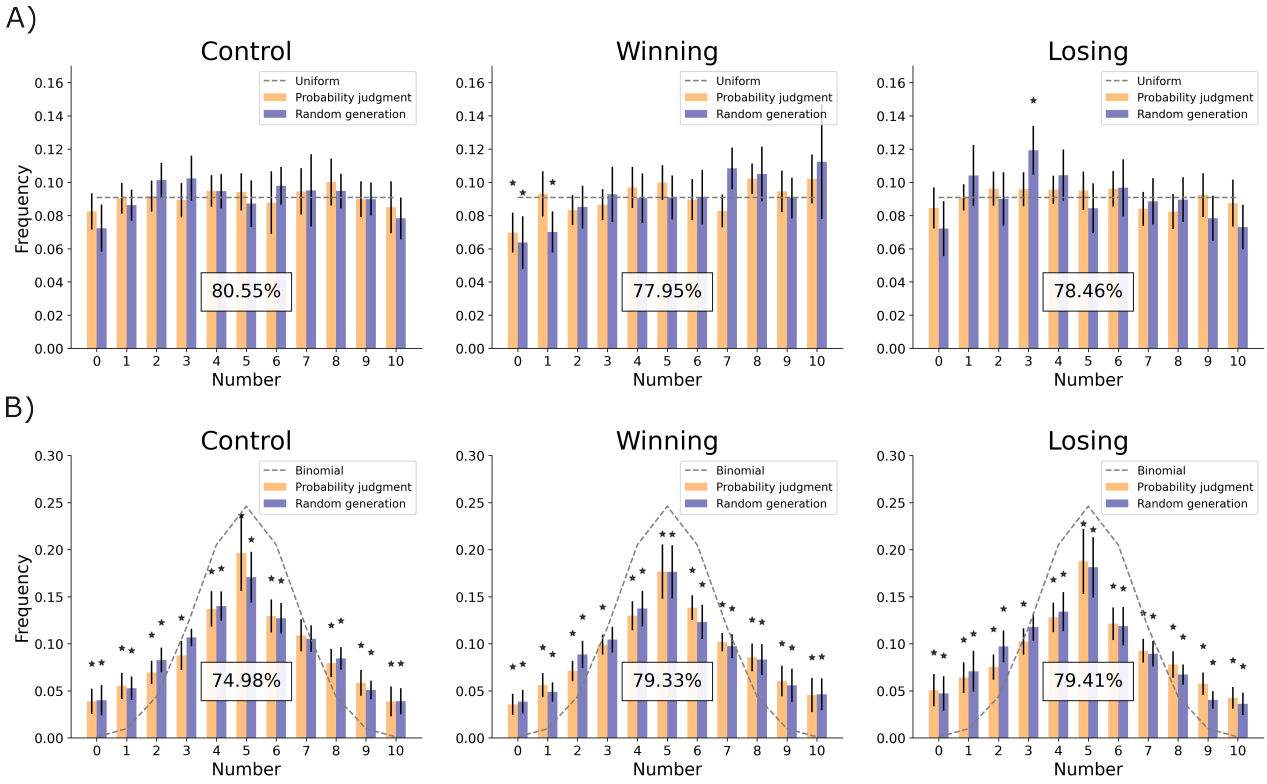

Figure 2: The comparison between the normalised probability judgment and the distributions of random generation in the control, winning and losing scenarios of (A) Experiment 1 and of (B) Experiment 2. Dashed lines indicate the theoretical distributions; black error bars represent the 95% confidence intervals; stars mark the numbers whose densities are significantly different from the theoretical distribution after a Holm-Bonferroni correction at the significance level of .05; and percentages are the overlapping coefficients between the probability judgment and random generation distributions.

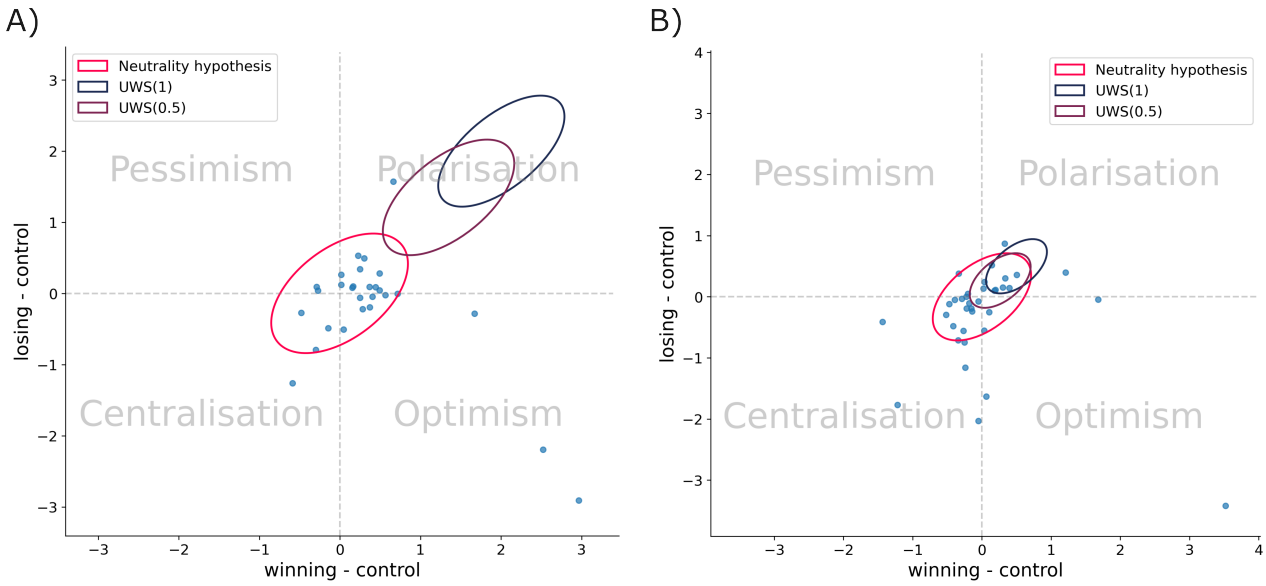

Figure 3: Scatter plots showing the relationship between the empirical winning-control mean difference and the empirical losing-control mean difference on the individual level in (A) Experiment 1 and in (B) Experiment 2. The red ellipse represents the 95% confidence region of the neutrality hypothesis; the deep blue ellipse represents the 95% confidence region of the UWS with the utility exponent of 1; the maroon ellipse represents the 95% confidence region of the UWS with the exponent of 0.5.

Table 1: Logarithm of the Bayes Factor (BF) for Each Model Compared with Model 0 in Experiment 1 and Experiment 2

index	model	lnBF in Exp 1	lnBF in Exp 2
0	$numbers = \beta_0 + u_{0,i}$	0	0
1	$numbers = \beta_0 + u_{0,i} + u_{1,i}scenario$	213.84	**335.39**
2	$numbers = \beta_0 + u_{0,i} + (\beta_1 + u_{1,i})scenario$	208.75	329.38
3	$numbers = \beta_0 + u_{0,i} + u_{1,i}scenario + \beta_2 equal$	210.29	332.78
4	$numbers = \beta_0 + u_{0,i} + (\beta_1 + u_{1,i})scenario + \beta_2 equal$	205.08	326.53
5	$numbers = \beta_0 + u_{0,i} + (\beta_1 + u_{1,i})scenario + \beta_2 equal + \beta_3 scenario \cdot equal$	198.94	319.93
6	$numbers = \beta_0 + u_{0,i} + u_{1,i}scenario + \beta_2 order$	210.17	331.52
7	$numbers = \beta_0 + u_{0,i} + (\beta_1 + u_{1,i})scenario + \beta_2 order$	204.69	325.44
8	$numbers = \beta_0 + u_{0,i} + (\beta_1 + u_{1,i})scenario + \beta_2 order + \beta_3 scenario \cdot order$	201.34	318.72
9	$\mathbf{numbers = \beta_0 + u_{0,i} + u_{1,i}scenario + \beta_2 st\_point}$	**229.42**	**334.38**
10	$numbers = \beta_0 + u_{0,i} + (\beta_1 + u_{1,i})scenario + \beta_2 st\_point$	224.34	328.50
11	$numbers = \beta_0 + u_{0,i} + (\beta_1 + u_{1,i})scenario + \beta_2 st\_point + \beta_3 scenario \cdot st\_point$	219.80	323.82

Note: The highest lnBFs and any with no substantial differences for each experiment are in bold. Additionally, the model with the highest lnBFs (considering equivalence) across the experiments is highlighted in bold. β represents fixed effects and u represents random effects. Four predictors were considered, namely the scenarios (*scenario*), participants' replies to the "equally likely" question (*equal*), the order of winning and losing scenarios (*order*), and whether the number is the first generated number in the sequence (*st_point*).

starting point of the winning scenario is numerically higher than the other two scenarios, the model comparison results in Table 1 argue against such an interaction.

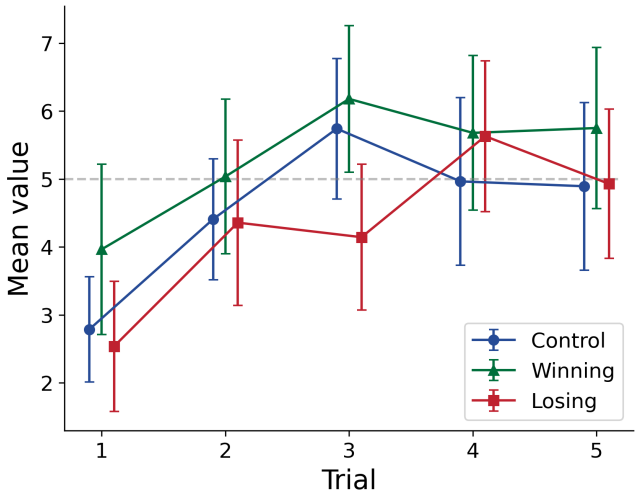

Figure 4: The means of the first five numbers generated in Experiment 1. The error bars represent the 95% confidence interval for each mean.

Experiment 2

While the results of Experiment 1 generally support the neutrality hypothesis, all three scenarios in this task involved a uniform distribution. As most risky events have unequal probabilities, Experiment 2 investigated whether these results hold for outcomes with unequal probabilities.

Participants

36 participants (29 women, 7 men; age: M = 22.9 years, SD = 3.7) took part in this experiment. Three additional participants were recruited but were excluded from the analysis since the length of their generated sequence was shorter than 80% of the theoretical length; and another one participant's data was excluded due to more than three values out of the range of 0 and 10. They received a show-up fee of £3, plus a potential bonus of £1, depending on the total points earned by the end of the experiment. The average payment was £3.50, and the experiment took about 40 minutes to complete.

Design, Materials and Procedure

The design and procedure were the same as Experiment 1, with the following exceptions. The materials were ten coins presented in a transparent box. Participants were asked to imagine shaking the box, observing the number of heads, and respond with that number. In the winning and losing scenarios, the number of heads represented the number of points that was won or lost, while in the control scenario, the number of heads had no meaning.

The order of the scenarios and the tempo of the flashing dot were identical to Experiment 1. In Experiment 2, participants produced a speed close to the theoretical one—an average of 148.26 numbers (SD = 6.19) after 4 minutes.

After the questionnaire, in the winning and losing scenarios, the experimenter shook the box in front of the participants, counted the number of heads, and converted it into points (one head equals one point) that the participants won or lost (see Footnote 1). Before the experiment ended, the experimenter calculated the total points (denoted as T). Then the participants were given a T/20 chance to win a £1 bonus or otherwise only received the show-up fee.

Results

Distribution of Responses Regarding participants' responses to the forced-choice question, the proportions of participants correctly selecting "no" were significantly higher than 50% in all three scenarios (Control: 25/36, $Z = 2.17, p = .015$; Winning: 26/36, $Z = 2.50, p = .006$; Losing: 25/36, $Z = 2.17, p = .015$) after a Holm-Bonferroni correction at a significance level of .05.

Figure 2B depicts the probability judgment and the distribution of random generation for each scenario compared with the binomial distribution. Notably, both the probability judgments and the distributions of random generation exhibited a significant departure from the expected binomial distribution, reflecting a flattened shape consistent with findings from prior research (Peterson, Ducharme, & Edwards, 1968; Teigen, 1974). The probability judgments and the random generation distributions were however again fairly similar to one another within participants, as measured by the overlapping coefficient (see Figure 2B).

Model Comparison of Random Generation Data The Bayesian linear mixed models and their lnBFs for Experiment 2 are shown in Table 1. Model 1 performed the best amongst these models. Like in Experiment 1, there was extreme evidence for individual differences in the effect of scenario (as Model 1 outperformed Model 0; $BF = 4.55 \times 10^{145}$) and there was also evidence against a fixed effect of scenario (comparing Model 2 with Model 1; $BF = 2.45 \times 10^{-3}$).

While in this experiment, Model 1 numerically outperformed Model 9, the Bayes factor in favor of Model 1 ($BF = 2.75$) showed it could only be classed as anecdotal evidence. Thus, it was inconclusive in this experiment whether there was a starting-point bias.

Individual Differences in the Effect of Scenarios We next explored the individual differences in the effect of scenario, employing the same MC method as in Experiment 1. As shown in Figure 3B, we found that 25 out of 36 participants' points (69%) fell in the 95% confidence region of the neutrality hypothesis, significantly higher than 50% ($p < .05$, 95% CI=[51.89%, 83.65%]). Such results imply that more than half of the participants were unaffected by the scenarios, i.e., the utilities.

We also compared the neutrality hypothesis with the UWS with two types of utility function defined above. The neutrality hypothesis outperformed the UWS models on 26 participants; the UWS with the exponent of 0.5 won on 8 participants; and the UWS with the exponent of 1 was supported by 2 participants.

General Discussion

Using the paradigm of random generation, we investigated the potential impact of outcomes' utilities on individuals' mental simulations with either a uniform or binomial payoff distribution. Across both experiments, for a majority of participants, the mean values of mental simulations appeared unbiased by utilities. These findings support the "neutrality" hypothesis outlined in Figure 1.

However, we also observed a utility-independent starting-point bias—a tendency to start from a lower number—though this bias was only evident in Experiment 1 and not replicated in Experiment 2. One plausible explanation for this absence may stem from the nature of the binomial distribution, which is more concentrated around the mean than the uniform distribution. This concentration could influence individuals' sampling process, starting from the distribution's central region. Consequently, detecting the bias becomes challenging due to its alignment with the probability influence. This discrepancy suggests that the bias is more likely to arise on conditions where all outcomes hold equal likelihood.

These discoveries hold theoretical implications for sampling-based models of probability judgment (e.g., Zhu et al., 2022, 2023) and risky choice (e.g., Busemeyer & Townsend, 1993; Busemeyer & Diederich, 2002; Roe et al., 2001; Lieder, Griffiths, & Hsu, 2018). These models assume that individuals make their decisions and judgments by sampling from specific distributions. Our findings indicate that such distributions should remain independent from outcome utilities, as in DFT but in contrast to models like UWS.

While primarily supporting the neutrality hypothesis, our results are concurrent with some hypothesises in Figure 1 while contradicting others. The observed individual differences align with the suggestion that only a subset of participants were pessimistic (Norem & Cantor, 1986), although we observed only one pessimistic participant in Experiment 2 and none in Experiment 1. The identified starting-point bias explains the findings supporting the centralisation hypothesis, where the risk-reward heuristic may be attributed to the initial recall of lower numbers. The difference between our results and those showing a polarization hypothesis could be because those studies used an experience paradigm in contrast to our description paradigm, and indeed polarisation seems to be driven by the encoding context rather than the choice context (Madan et al., 2021). Additionally, our findings might differ from studies on the optimism hypothesis due to instructional variations, as bias extents can vary based on different elicitation methods (Park et al., 2022).

There are however some limitations in our study, offering directions for future work. Firstly, the "random generation" instruction diverges from the common "prediction" instruction prevalent in this area of research; further research should thus investigate whether instructional changes influence random generation outcomes. Secondly, the differences among these outcomes might not be distinct enough to prompt a pronounced perception of utility disparity; employing distributions with more distinct outcomes in future research could therefore be valuable.

Acknowledgments

This research was funded by a European Research Council Consolidator Grant (817492-SAMPLING) awarded to ANS.

JQZ acknowledges the support of the NOMIS Foundation.

References

Baddeley, A. D. (1966). The capacity for generating information by randomization. *Quarterly Journal of Experimental Psychology*, *18*(2), 119–129.

Baddeley, A. D. (1998). Random generation and the executive control of working memory. *The Quarterly Journal of Experimental Psychology Section A*, *51*(4), 819–852.

Busemeyer, J. R., & Diederich, A. (2002). Survey of decision field theory. *Mathematical Social Sciences*, *43*(3), 345–370.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*(3), 432–459.

Castillo, L., León-Villagrá, P., Chater, N., & Sanborn, A. (2024). Explaining the flaws in human random generation as local sampling with momentum. *PLOS Computational Biology*, *20*(1), 1-24.

Glöckner, A., & Pachur, T. (2012). Cognitive models of risky choice: Parameter stability and predictive accuracy of prospect theory. *Cognition*, *123*(1), 21–32.

Hoffart, J. C., Rieskamp, J., & Dutilh, G. (2019). How environmental regularities affect people's information search in probability judgments from experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(2), 219–231.

Irwin, F. W. (1953). Stated expectations as functions of probability and desirability of outcomes. *Journal of Personality*, *21*, 329–335.

Krizan, Z., & Windschitl, P. D. (2007). The influence of outcome desirability on optimism. *Psychological Bulletin*, *133*(1), 95–121.

León-Villagrá, P., Castillo, L., Chater, N., & Sanborn, A. (2022). Eliciting human beliefs using random generation. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 44).

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

Lieder, F., Griffiths, T. L., Huys, Q. J. M., & Goodman, N. D. (2018). Empirical evidence for resource-rational anchoring and adjustment. *Psychonomic Bulletin & Review*, *25*(2), 775–784.

Ludvig, E. A., Madan, C. R., & Spetch, M. L. (2014). Extreme outcomes sway risky decisions from experience: Risky decisions and extreme outcomes. *Journal of Behavioral Decision Making*, *27*(2), 146–156.

Madan, C. R., Ludvig, E. A., & Spetch, M. L. (2014). Remembering the best and worst of times: Memories for extreme outcomes bias risky decisions. *Psychonomic Bulletin & Review*, *21*(3), 629–636.

Madan, C. R., Spetch, M. L., Machado, F. M. D. S., Mason, A., & Ludvig, E. A. (2021). Encoding context determines risky choice. *Psychological Science*, *32*(5), 743–754.

Marks, R. W. (1951). The effect of probability, pesirability, and "privilege" on the stated expectations of children. *Journal of Personality*, *19*(3), 332–351.

Nickerson, R. S. (2002). The production and perception of randomness. *Psychological Review*, *109*(2), 330–357.

Norem, J. K., & Cantor, N. (1986). Anticipatory and post hoc cushioning strategies: Optimism and defensive pessimism in "risky" situations. *Cognitive Therapy and Research*, *10*(3), 347–362.

Park, I., Windschitl, P. D., Miller, J. E., Smith, A. R., Stuart, J. O., & Biangmano, M. (2022). People express more bias in their predictions than in their likelihood judgments. *Journal of Experimental Psychology: General*, *152*(1), 45–59.

Pastore, M., & Calcagnì, A. (2019). Measuring distribution similarities between samples: A distribution-free overlapping index. *Frontiers in Psychology*, *10*.

Peterson, C. R., Ducharme, W. M., & Edwards, W. (1968). Sampling distributions and probability revisions. *Journal of Experimental Psychology*, *76*, 236–243.

Pleskac, T. J., & Hertwig, R. (2014). Ecologically rational choice and the structure of the environment. *Journal of Experimental Psychology: General*, *143*(5), 2000–2019.

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionst model of decision making. *Psychological review*, *108*(2), 370-392.

Teigen, K. H. (1974). Subjective sampling distributions and the additivity of estimates. *Scandinavian Journal of Psychology*, *15*(1), 50–55.

Weitzman, M. S. (1970). *Measures of overlap of income distributions of white and Negro families in the United States* (Vol. 22). US Bureau of the Census.

Zhu, J.-Q., León-Villagrá, P., Chater, N., & Sanborn, A. N. (2022). Understanding the structure of cognitive noise. *PLOS Computational Biology*, *18*(8), e1010312.

Zhu, J.-Q., Sundh, J., Spicer, J., Chater, N., & Sanborn, A. N. (2023). The Autocorrelated Bayesian Sampler: A rational process for probability judgments, estimates, confidence intervals, choices, confidence judgments, and response times. *Psychological Review*.