Cognitive Psychology 119 (2020) 101274

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Elucidating the differential impact of extreme-outcomes in perceptual and preferential choice

Yonatan Vanunu*, Jared M. Hotaling, Ben R. Newell

University of New South Wales, Australia

ARTICLE INFO

Keywords:
Risky-choice
Numerical averaging
Extreme outcomes
Time constraints
Selective-attention
Computational modeling

ABSTRACT

When making decisions in complex environments we must selectively sample and process information with respect to task demands. Previous studies have shown that this requirement can manifest in the influence that extreme outcomes (i.e. values at the edges of a distribution) have on judgment and choice. We elucidate this influence via a task in which participants are presented, briefly, with an array of numbers and have to make one of two judgments. In 'preferential' judgments where the participants' goal was to choose between a safe, known outcome, and an unknown outcome drawn from the array, extreme-outcomes had a greater influence on choice than mid-range outcomes, especially under shorter time-limits. In 'perceptual' judgments where the participants' goal was to estimate the arrays' average, the influence of the extremes was less pronounced. A novel cognitive process model captures these patterns via a two-step selective-sampling and integration mechanism. Together our results shed light on how task goals modulate sampling from complex environments, show how sampling determines choice, and highlight the conflicting conclusions that arise from applying statistical and cognitive models to data.

1. Introduction

Our visual system is often barraged with a vast range of stimuli. Thus, to be successful in our world we must selectively attend to the most relevant information. For example, an ordinary supermarket customer will not take the time to evaluate 25 assorted chocolates in an unfamiliar box before deciding whether to purchase it. Instead, the customer might selectively search for chocolates that he or she especially likes or dislikes, and decide accordingly. Hence, humans, as adaptive beings (Anderson, 1991; Payne, Bettman & Johnson, 1988, 1993), adjust their sampling and processing strategies in response to task demands, such as the goal they are attempting to achieve (e.g., getting the best chocolate and avoiding the worst) and the time available to make a decision. In the current study, we aim to test how decisions are modulated by task goals and time constraints in a decision paradigm that requires sampling from a complex visual array (i.e., an array of numbers and distractors). In addition, we present a cognitive model that explains how participants navigate the trade-offs between goals and time constraints.

The concept of selective sampling and processing in response to cognitive load or time constraints has a long history in the judgment and decision making literature. Much of this work has focused on rule-based heuristics in which some information is disregarded as a function of memory, validity, probability or outcome relevance (see Gigerenzer, 2004; Gigerenzer & Gaissmaier, 2011; Tversky, 1972). For instance, Gigerenzer and Goldstein (1999) proposed the Take-The-Best heuristic that determines choice according to the prediction of the best or most valid cue in a given environment, while ignoring all other information. In the context

* Corresponding author at: School of Psychology, University of New South Wales, Sydney 2052, Australia.
E-mail address: yyv1984@gmail.com (Y. Vanunu).

https://doi.org/10.1016/j.cogpsych.2020.101274
Received 23 May 2019; Received in revised form 5 January 2020; Accepted 7 January 2020

of experience-based decisions, it has been suggested that people rely on small or partial samples from memory when making choices (Hau, Pleskac, Kiefer, & Hertwig, 2008; Hertwig & Pleskac, 2008, 2010), and that these small samples can sometimes be unrepresentative of the population-level properties of an outcome-distribution (Lieder, Griffiths, & Hsu, 2018).

In the current study, we focus on how people adapt to environmental constraints by investigating the influence of 'extreme items' in a distribution. We define extremity as outcomes at the edges of a distribution. For example, in the sample [11, 23, 45, 37, 42, 89] – the most extreme outcomes are [11] and [89]. We choose to focus on this aspect of choice because of recent, seemingly conflicting findings regarding the impact of extreme outcomes on people's decisions. Specifically, studies examining risky-choice have shown that extreme items receive more attention (Kunar, Watson, Tsetsos, & Chater, 2017; Pleskac, Yu, Hopwood, & Liu, 2019; Tsetsos et al., 2016; Tsetsos, Chater, & Usher, 2012; Vanunu, Pachur, & Usher, 2019; Zeigenfuse, Pleskac, & Liu, 2014) and are better remembered (Lieder et al., 2018; Ludvig & Spetch, 2011; Ludvig, Madan, & Spetch, 2014; Ludvig, Madan, McMillan, Xu, & Spetch, 2018; Madan, Ludvig, & Spetch, 2014) than mid-range items and as such, have a greater impact on choice. In contrast, recent evidence suggests that extreme items receive less attention in numerical averaging tasks and therefore have a smaller impact on choice than those in the middle of the distribution (Vandormael, Castañón, Balaguer, Li, & Summerfield, 2017; Li, Castañón, Solomon, Vandormael, & Summerfield, 2017; see also De Gardelle & Summerfield, 2011). We attempt to identify the source of this discrepancy and its implications for choice.

Although previous studies have focused on the impact of presentation format (i.e. sequential vs simultaneous stimulus presentation) on risky choice and numerical averaging (Brusovansky, Vanunu, & Usher, 2019; Vanunu et al., 2019), in the current study we focus on a different feature that sets the two tasks apart – the participant's goal. In numerical averaging tasks, participants are required to produce an estimation of the objective average for the numbers presented in an array, and are rewarded according to their accuracy. By contrast, in risky-choices, people choose according to their subjective preference (e.g., risk seeking) and are rewarded with the probabilistic outcome of their choice. In the following, we will refer to averaging paradigms as *perceptual tasks* and to risky-choices as *preferential tasks* (following Dutilh & Rieskamp, 2016)[1].

We suggest that extreme items might play different roles in determining performance in these two tasks, especially when time constraints are imposed. Ideally, the optimal strategy should be to sample and process all items equally before making a choice. However, processing information and making a decision is effortful and time consuming (Ariely & Zakay, 2001). In perceptual tasks - where one is presented with an array of numbers and is required to quickly produce an estimation of the stimulus' average – people might allocate more attention to the mid-range items as they are similar to the true mean and therefore might serve this task goal better. Under such conditions it is reasonable to expect that the extremes will be the first to "drop" from processing, similar to the practice of excluding outliers as "noise" in descriptive statistics. Vandormael et al. (2017) have named this mechanism *robust averaging* and highlighted its predictive power in a task where participants had to estimate the average of an array of numbers under various time constraints. A follow-up study found that robust averaging can enhance performance in the presence of neural-computational noise (Li et al., 2017).

In contrast, in preferential tasks where a participant might be motivated to "win the best and avoid the worst", more attention might be placed on the extreme items in a distribution as they are the best and the worst. This selective mechanism should be enhanced when cognitive resources are scarce (Payne et al, 1988, 1993) and indeed, most studies that have found a greater influence of extreme items on risky-choice have tested complex choice environments that impose high demands on cognitive capacity (e.g., Kunar et al., 2017; Pleskac et al., 2019; Tsetsos et al., 2012, 2016; Vanunu et al., 2019; Zeigenfuse et al., 2014; though see Madan, Spetch, & Ludvig, 2015 for a counterexample).

We aim to answer two questions: (a) what is the nature of the influence of extreme items on performance in perceptual and preferential tasks, and (b) what do these differences imply in terms of the cognitive mechanisms subserving performance in the two tasks? Toward this goal, in two experiments we tested a novel behavioural paradigm in which we manipulate task goal within participants and time constraints between participants while holding presentation format constant (for details about the experimental time-line see S1 in the supplement material). To manipulate task goal we created perceptual and preferential versions of a similar task. In the perceptual task, an array of numbers and letters was displayed for a short period and participants were asked to decide whether the average of these numbers was larger than a reference value (R). The letters simply acted as distractors to make the visual search of the numbers more taxing. Participants were rewarded for the accuracy of their choices. In the preferential task, the same procedure was implemented, but the numbers in the array were described as monetary payoffs, and participants had to decide whether they wanted to take a gamble to win one of those payoffs or to take a certain payoff (i.e. R; see Fig. 1). Choosing the certain payoff ensured that amount as a reward, whereas choosing the array led to an equal-probability draw from the values in the array (i.e., one in which there was an equal probability of receiving the best, worst, or any other value). As such, participants observed the same stimuli in the two tasks but had different goals and different incentive structures.

We also manipulated the distribution of values in the arrays to explore further the impact of extreme items. Specifically, in Experiment 1 we included a within-subjects manipulation of variance in which half of the presented arrays had high variance and half had low variance. In Experiment 2 we fixed the variance but manipulated the skew of the array within-subjects to test the impact of low extremes (left skew array) and high extremes (right skew array) on separate trials. Finally, time constraints were implemented between groups by manipulating how long the arrays appeared on the screen.

[1] We acknowledge that choices in the perceptual tasks are also based on the participants' subjective impression, and that the preference in the preferential task is formed from the perceptual input of the stimuli. However, we find these precedent definitions are sufficient to illustrate the core differences between the two tasks' goals (see Dutilh & Rieskamp, 2016).

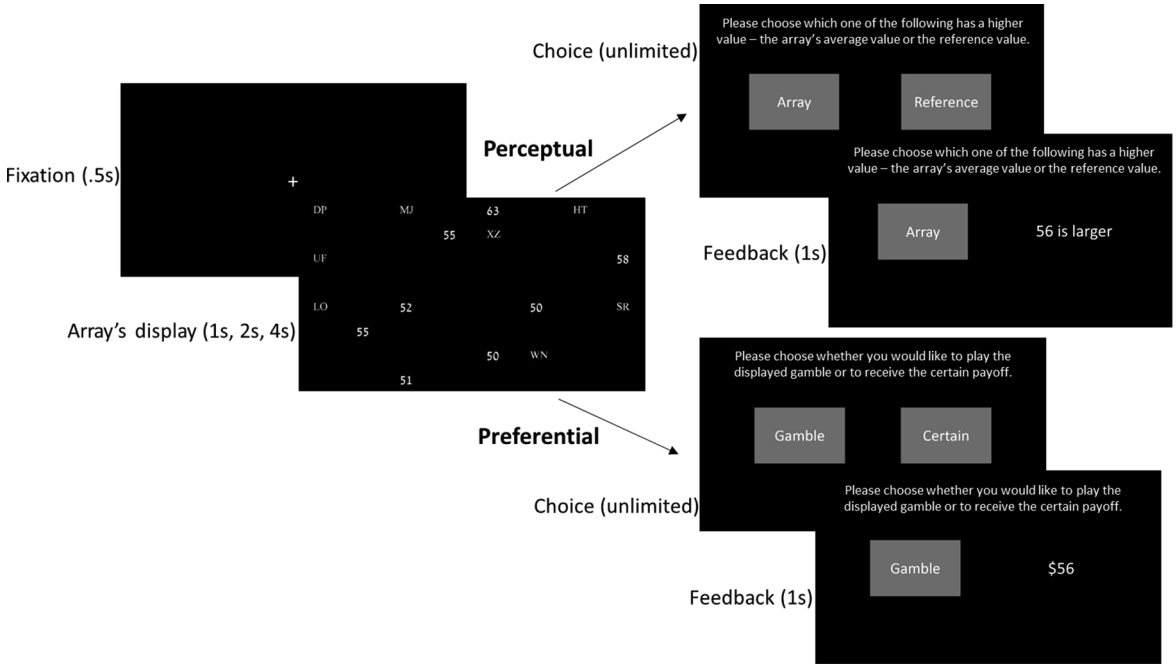

Fig. 1. Example trial display from the perceptual and the preferential tasks in Exp. 1. In both tasks, participants could view the numbers on the screen for 1s, 2s, or 4s, and make a decision accordingly. Each block started with a display of the reference value/certain payoff, but this is not illustrated in the figure. The values in the brackets indicate the screen's display-time.

Our basic predictions follow from prior work that has observed *robust averaging* in perceptual tasks similar to ours (e.g., Vandormael et al., 2017) and *risk-seeking* in preferential tasks (e.g., Vanunu et al., 2019). As such we expect variance to have more influence in the preferential than the perceptual task, because in the risky-choice context of the former, participants might be drawn to seek the best/avoid the worst outcome, but such a tendency should not be invoked by the goals/incentives of the perceptual task. We also expected that difference to be amplified under shorter time constraints due to increased attention to the task-relevant extremes when the opportunity to sample is limited (e.g., Payne et al, 1988, 1993).

Going beyond these behavioural predictions, following the experiments, we present a modelling analysis that outlines a single cognitive procedure that can describe how people allocate attention to different items as a function of their extremity. Our basic conjecture is that participants sample information 'online', strategically attending to the most relevant items. A probabilistic sampling function determines the probability with which values in the array will be sampled and in turn integrated to drive choice. This sampling probability is guided by the participant's goal in the given task. Critically, the proposed sampling function differs from common attentional-weighting functions in the literature (e.g., Spitzer, Waschke, & Summerfield, 2017; Tsetsos et al., 2012, 2016; Vanunu et al., 2019; Zeigenfuse et al., 2014) because it does not assume that all items are processed and weighted in choice. Instead, it posits that item magnitudes are preserved (i.e. "unweighted"), but that sampling is probabilistic and incomplete.

In our modelling analysis we analyse these sampling functions and examine two plausible procedures for integrating items after they have been sampled: *averaging* and *counting*. The first entails estimating and comparing the arrays' average to the reference value. This can be optimal for both tasks as it corresponds to the goal of the perceptual task and allows for the maximization of expected value in the preferential task. The counting procedure might be implemented in response to the probabilistic outcome design of the preferential task (i.e. a random draw from the array). Here participants compare the number of items that are above and below the reference value, and choose the option supported by the majority of items (see Thorngate, 1980). In other words, the counting procedure chooses based on the probability of receiving an outcome that is larger than the reference number/certain payoff. Our analysis allows us to evaluate the evidence for these two types of procedures, and to demonstrate how task-type, goals, time constraints and the properties of the array's distribution affect choice.

To foreshadow our results, across two experiments and the modelling analysis we found evidence that participants engaged in strategic and incomplete information sampling. In particular, they were more likely to sample extreme items, especially in the preferential task and under very short time constraints. Moreover, although participants were adaptive and adjusted integration according to task-demands, the incomplete sampling strategy induced some participants to take more risks which, in some cases, diverged choice away from maximization.

2. Experiment 1

2.1. Method

2.1.1. Participants

120 first-year students from the University of New South Wales (82 females, age: 17–37 years, M = 19.41) participated in Experiment 1[2], in exchange for course credits and an incentivization bonus. Twelve participants were excluded from the analysis because they did not attend the second experimental session, and one participant was excluded for making the same choice on every trial.

2.1.2. Design & materials

We used a 3 (Display-Time: 1 s, 2 s, 4 s) × 2 (Task-Type: Perceptual, Preferential) × 3 (Mean Differences: Below, Above, Approximately equal) × 2 (Variance: Low, High) design. The first factor was implemented between-subjects and the remaining factors were within-subjects. Task-Type was blocked so that participants completed one session of either the perceptual or preferential task and then a week later completed the other version. All other factors were manipulated within sessions.

Array properties: Five reference values (R) were drawn randomly from a uniform distribution, $U(50,60)$, one for each experimental block. For each trial, an array of eight two-digit numbers and eight two-letter strings were randomly placed on screen within an invisible 8 × 8 grid. The grid divided the display (1920 × 1080 pixels resolution) into 64 equally sized rectangles (240 × 135 pixels each; see Fig. 1). No stimuli were displayed within the central four grid locations. Three hundred sets of two digits numbers (i.e. ranged 10–99) were randomly drawn from one of six Gaussian distributions. The distributions had either high variance (SD = 15) or low variance (SD = 5). Their mean (μ) was set according to R, and was either below (μ = R-3), above (μ = R + 3) or approximately equal ($-1 \leq \mu\text{-}R \leq 1$). The number of trials was equally divided among the six conditions of variance and mean differences. The letter strings were randomly selected from the English alphabet and served as distractors.

2.1.3. Procedure

Across two sessions separated by a week, each participant carried out both the preferential and the perceptual task, with session order counterbalanced across participants. At the beginning of each session, participants were instructed that they were about to see an array of numbers and letters and to make a decision accordingly. In the perceptual task, participants were instructed to compare the average value of the numbers in the array to a reference value, and choose the one with the larger value. They were also informed that at the end of this session, eight trials would be randomly selected and each correct answer among these trials would grant them with $1. In the preferential task, the numbers in the array were described as monetary payoffs, and participants were instructed to choose between two options. They could receive the reference amount, R, with certainty (certain payoff), or they could take a gamble and receive one of the eight amounts displayed in the array, with equal probability. Participants were also informed that at the end of this session, a random trial would be selected by the computer, and 10% of that trial's outcome would be granted to them as a reward. Finally, in both tasks, participants were instructed that the letters in the array had no meaning. Following the instructions in both sessions, a short practice phase was given where an example reference value was initially displayed, followed by six example trials that corresponded to one of the six variance × mean differences conditions above.

In the experimental-trials, each block started with a display of the reference value and participants were instructed to remember it well since this value would apply for all 60 trials in the block. On each trial, the array was displayed for 1 s, 2 s or 4 s, depending on the between-subject display-time condition. Following the removal of the array, a choice-screen was presented (Fig. 1). Participants were not allowed to make a decision before the choice-screen display, but no additional constraints on the decision time itself were imposed. After a decision was made, feedback was given in two forms. In the preferential task the outcome of the participant's choice was displayed on the screen (i.e. the value of the certain payoff or a random draw from the array), and a high or low pitch tone was played to indicate whether they received the higher or lower payoff between the two alternatives, respectively. In the perceptual task the feedback corresponded to the objective performance criterion, presenting a binary outcome of the answer (e.g., "array is larger" or "57 is smaller"; see Fig. 1), and a high or low pitch tone was played to indicate whether they were correct or not. In both the preferential and perceptual sessions, participants completed five experimental blocks of 60 trials, where each 10 trials corresponded with one of the variance × mean differences conditions, with trial order randomized across block.

2.2. Results

We analysed two aspects of the data. Our first analysis focusses on the rate at which participants maximized performance (i.e. choosing the option with the larger expected value/average) by calculating the proportion of trials on which the alternative with a higher average was chosen[3]. This measure allows us to see whether expected effects of mean differences (Below/Above >

[2] See S1 in the supplement material for an explanation of our sampling plan and the experimental time-line. The supplement also contains details of preregistered predictions for our experiments. Note that the sample size was set to exceed the sample sizes in previous studies (Vandormael et al., 2017; Tsetsos et al., 2012; Vanunu et al., 2019).

[3] In R ≈ μ condition, the difference between R and μ on most trials was smaller than one, but nonzero. These differences enabled us to determine the better option when calculating maximization rates. Cases where R and μ have identical values were treated as trials that cannot be maximized by averaging and therefore, were analyzed as missing data in the maximization rates analysis.

approximately equal), display-time (Longer > Shorter), and variance (Low > High) are obtained. This analysis can be interpreted as a basic check that participants are paying attention to the task. Our second analysis is the one of primary interest. Here we examine the impact of the same variables on the proportion of array choices (i.e. p[array]). This analysis allows us to focus directly on the way in which the array's variance might interact with task-type and display-time to produce the differential impact of extreme values on choice. For the analyses we applied a logit mixed-effects model with either maximization or p(array) as the dependent variable. We pooled the data across participants and set Display-time, Task-type, Variance and Mean differences as fixed effects, and Participant ID as a random intercept effect. To test the significance of each component in the model, we applied a likelihood-ratio chi-squared test between the mixed effects model and a similar model that is missing the components of interest, using the 'car' package in R (Fox et al., 2012).

2.2.1. Maximization

Maximization: The analysis revealed the expected effects, showing that participants maximized more when the array was displayed on screen for longer ($M_{1s} = 0.60$, SD = 0.06; $M_{2s} = 0.66$, SD = 0.05; $M_{4s} = 0.71$, SD = 0.04; $\chi^2(2) = 76.12$, $p < .001$), when the array's variance was low ($M_{lowV} = 0.71$, SD = 0.08; $M_{highV} = 0.61$, SD = 0.05; $\chi^2(1) = 692.36$, $p < .001$), and when the difference between R and the mean of the array was larger ($M_{below} = 0.72$, SD = 0.10; $M_{above} = 0.73$, SD = 0.11; $M_{equal} = 0.52$, SD = 0.04; $\chi^2(2) = 2365.12$, $p < .001$). Interestingly, we also found that participants tended to maximize more (i.e., be more accurate) in the perceptual task than the preferential task ($M_{perc.} = 0.67$, SD = 0.07; $M_{pref.} = 0.65$, SD = 0.07; $\chi^2(1) = 31.97$, $p < .001$). This suggests that people used a different (sub-optimal) strategy in the preferential task – a tendency that is further elucidated in the analysis of the proportion of array choices. A number of the two-way and three-way interactions were also statistically significant, but as they are not of direct theoretical interest we do not discuss them here. A full report (with a figure showing maximizing rates) can be found in S2 in the Supplementary Materials.

2.2.2. Proportion of array choices

Fig. 2 (A; top panels) shows the proportion of array choices as a function of task-type, display-time and variance, but collapsed across mean differences (for a full report see S2 in the supplement material). The figure shows consistently larger preference for high variance arrays than low variance arrays in the preferential task, but indifference in the perceptual task across all display-time groups. This pattern suggests a tendency to be more influenced by the high extremes present in the high variance array when participants are making risky-choices than when they are asked to estimate an average.

We found a main effect of variance with a higher choice proportion for high variance than low variance arrays ($M_{highV} = 0.52$, SD = 0.08 vs. $M_{lowV} = 0.49$, SD = 0.07; $\chi^2(1) = 45.51$, $p < .001$). However, more importantly and in line with our predictions, a two-way interaction between task-type and variance was found ($\chi^2(1) = 31.73$, $p < .001$), indicating that participants were risk seeking and chose the array more often in the preferential task when the array's variance was high than when it was low ($M_{highV} = 0.53$, SD = 0.12 vs. $M_{lowV} = 0.48$, SD = 0.10; $\chi^2(1) = 73.81$, $p < .001$). By contrast, participants were indifferent to variance in the perceptual task ($M_{high} = 0.52$, SD = 0.07 vs. $M_{low} = 0.51$, SD = 0.07; $\chi^2(1) = 0.92$, $p = .34$). Surprisingly, no main effect of display-time nor interactions with the other variables were found.

2.3. Discussion

The interaction between variance and task goal supports our predictions for the effect of task goal on choice, and the role of extreme items in achieving these goals. On the one hand, the effect of variance (i.e. risk seeking) found in the preferential task suggests that high extreme payoffs were more likely to be sampled than mid-range or low extreme items because over-sampling these items would have increased the attractiveness of high variance arrays (e.g. Tsetsos et al., 2012). On the other hand, indifference between high and low variance arrays in the perceptual task suggests that high extreme items were not oversampled in this task.

In order to test the impact of the high and low extremes on separate trials, in Experiment 2 we fixed the variance but manipulated the skew of the arrays to produce arrays with only high extremes (i.e. a right skew array) or only low extremes (i.e. a left skew array). We predicted that while high extremes would have the greatest impact on choice, participants would also oversample low extremes when they appear in separate trials with only mid-range items. We also expected these effects to grow with greater time constraints.

To make the two tasks more similar (in terms of feedback) we changed the perceptual design to include a probabilistic outcome, as in the preferential task. That is, we asked participants to estimate which value is larger: a reference number or a random draw from the array. This equates the two tasks in terms of feedback, but maintains the difference in incentive schemes because participants in the perceptual task were still rewarded based on accuracy (i.e. whether or not the sampled outcome was larger than the reference number).

Following these modifications, we predicted that the new probabilistic design of the perceptual task would promote a counting procedure. This procedure is predicted to induce preference for arrays with a left skew over arrays with a right skew because in the left-skew arrays there are more items above the reference number and in the right skew array there are more items below the reference number. We also expected to find opposite choice patterns between left and right skew arrays as a function of task goal.

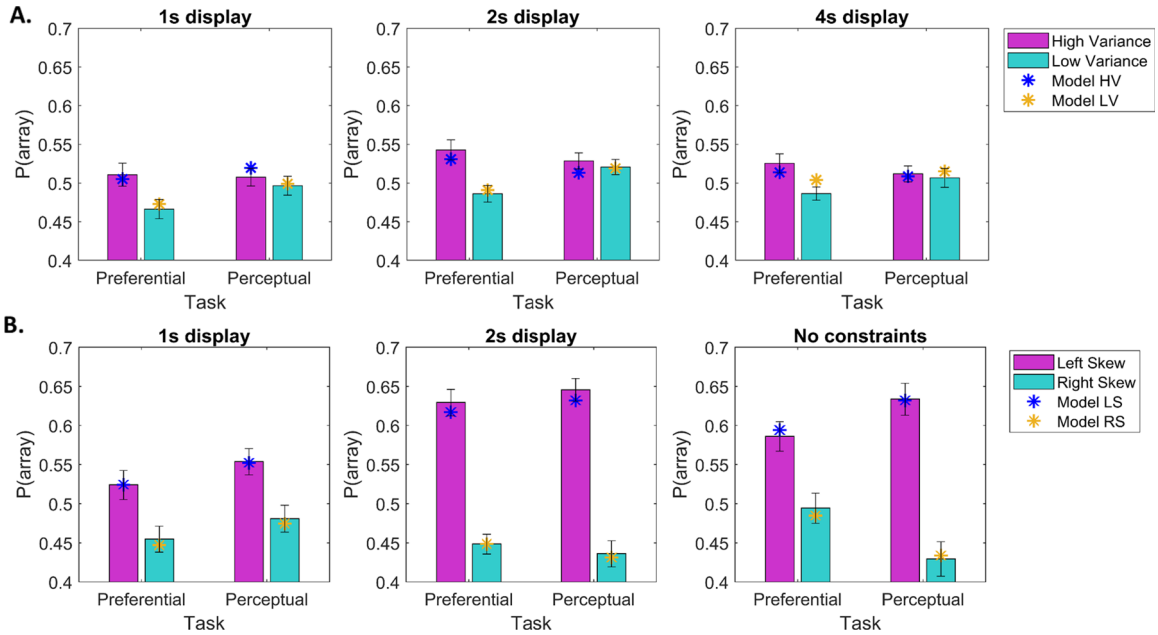

Fig. 2. The average proportions of choosing the array between display-time groups, task-types, low/high variance arrays in Exp. 1 (panel A) and left/right skew arrays in Exp. 2 (panel B). In all panels the data are collapsed across mean differences. The asterisks illustrate the SSIM's predictions according to the participants' best fitting model. Error bars correspond to within-subjects standard errors.

Specifically, since we expected the extremes to have a greater impact on choice in the preferential task, we predicted that arrays with a right skew, which include a very high outcome but only moderately low outcomes, would be chosen more often than arrays with a left skew, which include a very low outcome but only moderately high outcomes.

3. Experiment 2

3.1. Method

3.1.1. Participants

Ninety first-year students from the University of New South Wales (60 females, age: 17–48 years, M = 19.61) participated in Experiment 2, in exchange for course credits and an incentivization bonus[4]. Two students were excluded from the analysis since they did not attend the second session.

3.1.2. Design & materials

We used a 3 (Display-Time: 1 s, 2 s, No constraints) × 2 (Task-Type: Perceptual, Preferential) × 3 (Mean differences: Below, Above, Approximately equal) × 2 (Skew: Left, Right) design. The first factor was implemented between-subjects and the remaining factors were within-subjects

Array properties: The stimuli in Experiment 2 were identical to Experiment 1 with the following exceptions: the display screen was divided into an invisible 10 × 10 grid of 100 equally sized rectangles (192 × 108 pixels each); with no stimuli displayed within the central four grid locations or in the first or last column or row of the grid. Two hundred and eighty sets of two digits numbers (eight numbers in each set) were randomly drawn from normal distributions with high variance (SD = 15), and with means that were either below ($\mu = R-3$), above ($\mu = R + 3$) or approximately equal to R ($-1 \leq \mu-R \leq 1$). However, in Experiment 2 trials with approximately no mean differences constituted half of the trials[5]. To implement a skew manipulation, we measured the skewness of the produced arrays using the Fisher-Pearson standardized third moment coefficient (see Doane & Seward, 2011), and divided them into right skew, left skew, and no skew groups[6].

[4] See S1 in the supplement material for details of our sampling plan and the experimental time-line.

[5] We increased the number of $R \approx \mu$ trials since an incomplete sampling strategy might operate as a tiebreaker when the distribution of items below and above the R is more even - i.e. a selective sampling strategy should have a greater impact on choice in $R \approx \mu$ trials.

[6] We excluded "no skew" trials from the statistical analysis as they imposed complications with an extra level, but had little theoretical contribution. However, all conditions were included in the computational modeling analysis. A report on the no skew condition can be found in Figure S5 in the supplement material.

3.1.3. Procedure

The overall procedure in Experiment 2 was similar to the procedure in Experiment 1, with one session for each task separated by a week. Instructions and feedback in the perceptual task mirrored the preferential task and used the same probabilistic outcome design. Participants in the perceptual condition were asked to estimate which one of the following would have a larger value: a random draw from the displayed array or a reference value. To make their choice, participants pressed one of two buttons – one for the array and one for the reference value - which were displayed at the bottom of the screen throughout the trial. This modification to the procedure of Experiment 1 allowed participants to respond while still viewing the array (as opposed to when the choice screen appeared in Experiment 1 – see Fig. 1). Stimuli were displayed on screen for either 1 s, 2 s, or until a choice was made (no constraints), depending on the Display-Time condition. Following choice, feedback with the outcome of their choice was displayed on screen (a random draw or R), and a high or low tone indicated whether they received the higher or lower outcome between the two alternatives. Each task began with a short practice phase with eight example trials and an experimental phase with five blocks of 56 trials.

3.2. Results

3.2.1. Maximization

Due to the introduction of the probabilistic outcome design in Experiment 2, we calculated maximization rates based on the number of items above the reference value (i.e. consistent with the counting procedure). If the majority of items were above R, the correct response was to choose the array, otherwise the correct response was to choose R[7]. Nonetheless, it is important to note that while counting is the optimal strategy for choice under the new perceptual design, both counting and averaging are appropriate in the preferential task, because the former maximizes expected value (EV), and the latter maximizes the probability of winning. We applied the same mixed-effect logistic regression model as in Experiment 1, but with skew replacing variance. Similar to Experiment 1, we found the lowest maximization rates under a 1 s display, but surprisingly we found that participants maximized more when the array was displayed on screen for 2 s than with no constraints on the display ($M_{1s} = 0.57$, SD = 0.06; $M_{2s} = 0.67$, SD = 0.08; $M_{no} = 0.64$, SD = 0.13; $\chi^2(2) = 12.68$, $p < .005$). Participants also maximized more when the difference between R and the mean of the array's values was larger ($M_{below} = 0.62$, SD = 0.12; $M_{above} = 0.64$, SD = 0.12; $M_{equal} = 0.61$, SD = 0.11; $\chi^2(2) = 29.42$, $p < .001$); more when the array's skew was left than when it was right ($M_{leftS} = 0.65$, SD = 0.13; $M_{rightS} = 0.60$, SD = 0.12; $\chi^2(1) = 48.38$, $p < .001$); and more in the perceptual task than the preferential task ($M_{perc.} = 0.64$, SD = 0.13; $M_{pref.} = 0.61$, SD = 0.11; $\chi^2(1) = 28.61$, $p < .001$; for the full report see S3 in the Supplementary Material).

3.2.2. Proportion of array choices

Fig. 2 (B; bottom panels) shows the proportion of array choices as a function of task-type, display-time and skew but collapsed across mean differences (for a full report see S3 in the Supplemental Material). The figure shows a clear dominance of the left skew arrays over the right skew arrays across all conditions. However, this dominance seems to diminish in the preferential task under no time constraints and in both tasks under a 1 s display. This implies that under these conditions the display of high extreme items (right skew condition) increased the array's attractiveness, and the display of low extreme items (left skew condition) decreased the array's attractiveness.

In support of these observations, we found a main effect of skew ($M_{left} = 0.60$, SD = 0.11; $M_{right} = 0.46$, SD = 0.10; $\chi^2(1) = 613.82$, $p < .001$), indicating that participants chose the left-skew arrays more often than the right-skew arrays. The two-way interactions between task-type and skew ($\chi^2(1) = 17.90$, $p < .001$) and display-time and skew ($\chi^2(2) = 83.24$, $p < .001$) were statistically significant, indicating that left skew arrays were chosen more often than right skew arrays in the perceptual task than in the preferential task; and under shorter display-time, these differences diminished. Most importantly, a three-way interaction between all three variables was found (i.e. display-time × task-type × skew; $\chi^2(2) = 16.43$, $p < .001$), supporting our observation of a diminished dominance for a left skew display over a right skew display as a function of the manipulated variables. That is, for both tasks under 1 s display and for the preferential task under no constraints, choice proportion for left skew arrays decreased and choice proportion for right skew arrays increased, implying that both high and low extremes had a greater impact on choice when displayed in separate arrays (i.e. left and right skew display).

3.3. Discussion

In line with our predictions, we found a greater preference for a left skew display than a right skew display in the perceptual task. This is consistent with the idea that the probabilistic outcome design in Experiment 2 led participants to use a counting procedure that favours a left-skew array (because there are more numbers above the reference value in a left-skew array; see Fig. S6 in supplement material for the distribution of the count, average and min. & max. values between the skew conditions). However, this preference for left-skew arrays was also found for the preferential task – contrary to our prediction that right-skew arrays would dominate in the preferential task because of the presence of the high extreme values in these arrays. The persistent appeal of the left-

[7] Cases where equal amount of items in the array were below and above the R were treated as trials that cannot be maximized by counting and therefore, were analyzed as missing data in the maximization rates analysis.

skew arrays suggests that a counting procedure was also employed in the preferential task. Partial support for a limited influence of the right-skew arrays was found in the three-way interaction which implied that in the preferential task and under shorter time constraints, both the high and the low extremes had a greater impact on choice than the mid-range items. This diminished the difference in risky-choice proportions between left and right skew displays in these conditions (indicated by larger min. & max. values in a right-skew than a left-skew display, see Fig. S6 in supplement material), though the impact was not sufficient to reverse preferences completely. Surprisingly, the difference between left and right skew display was smaller under no time constraints than in the 2 s condition, implying that participants did not utilize additional time to sample the entire display. In support of this interpretation, participants exhibited similar response times (RT) in the 2 s and no constraint conditions (median RT values: $RT_{1s} = 1.64$ s, $RT_{2s} = 2.63$ s and $RT_{no} = 2.66$ s; for full report see Table S5 in the supplement material).

Looking together at the behavioural findings in Experiments 1 and 2, we find clear evidence that choices were affected by the presence of extreme items in the displayed arrays, and that these effects interacted with task-goal and display-time. Specifically, in the preferential task and under shorter time constraint, an increase in choice proportion for high variance arrays in Experiment 1 and for right skew arrays in Experiment 2, implied that participants were more likely to sample the high extreme items from the arrays. Additionally, choice proportion for the left skew arrays in Experiment 2 decreased under the same conditions, suggesting that participants were more likely to sample the low extreme items. Combined, these results point to a flexibility in people's decision making whereby they selectively sample only the most strategically relevant information on display.

4. The selective sampling and integration model

We now present a cognitive model to account for the behaviour observed in Experiments 1 and 2. The *Selective Sampling and Integration Model* (SSIM) describes a decision process comprised of two steps: *selective sampling* – when information is collected selectively in accordance with task goals and cognitive constraints – and *integration* – when information is combined through one of two procedures, *averaging* or *counting*. Importantly, the model diverges from earlier work (e.g. Vandormael et al., 2017) by paying close attention to the process by which people collect information from a display. By comparing the SSIM model to an earlier statistical model (see Section 5) we demonstrate how our approach provides new and greater insight into the cognitive mechanisms underlying observed behavior.

4.1. Selective sampling

The SSIM posits that people may not sample all of the items displayed in an array, but may instead selectively sample items based on the goal they are trying to achieve and the cognitive resources they have available. Note that 'sampling' refers to the transferring of a selected item into integration. This means that participants may perceive an item, but fail to sample it for integration. We formalize this selective sampling process with the *Probabilistic Sampling Function* (PSF), which defines the probability that each item is sampled. We assume that items are sampled according to their extremity, defined as the item's rank within the displayed array (e.g. the smallest value item would have rank 1 and the largest value item would have rank 8)[8].

Fig. 3A illustrates some possible PSFs and how their shapes determines an individual's sampling strategy. A U-shape function (blue curve in Fig. 3A) indicates a higher probability of sampling the highest and lowest ranks, which would result in the extremes having a greater impact on choice. An inverse U-shape function (red curve in Fig. 3A) indicates the opposite: mid-range items are more likely to be sampled and will have the greatest impact on choice. Asymmetric PSFs are also possible. For example, the yellow and purple curves in Fig. 3A represent policies for selectively sampling high and low extremes, respectively. Finally, the area under the curve indicates how many items will be sampled into processing on each trial (i.e. sample size).

Mathematically, the PSF is an elaboration of a quadratic function in two steps. First, we define the shape of the function in Eq. (1) with the product q(X). Then, we transform q(X) into a probability vector p(X) of sampling each item, X_i, of rank i in an array (Eq. (2)):

$$q(X_i) = \alpha \times (z_i - \theta)^2 \tag{1}$$

$$p(X) = q(X) - \min(q[X]) + \beta \tag{2}$$

Here z is a rescaling of ranks 1–8 to evenly spaced values between -1 and 1. Three parameters define the shape of the function. First, α determines the curvature of the function, where $\alpha > 0$ produces a U-shape function (favoring the extremes) and $\alpha < 0$ produces an inverse-U shape function (favoring the mid-range items). θ determines the symmetry of the sampling function, where $\theta > 0$ indicates higher probabilities for higher ranks and a $\theta < 0$ indicates higher probabilities for lower ranks. Next, a normalization function is applied by subtracting the minimum point of q(X) from all values in q(X), followed by adding β to all values to determine the area under the curve and the sample size. $\beta = 1$ indicates that all items in the display are sampled, and $\beta = -1$ indicates that no items in the display are sampled. All parameters are bounded between -1 and 1. For example, the parameters values [α, θ, β] that produced the example PSFs in Fig. 3A are [0.5, 0, 0.5] for the blue curve, [-0.5, 0, 0.4] for the red curve, [0.15, 1, 0.4] for the yellow curve and [0.15, -1, 0.4] for the purple curve.

[8] An alternative model in which extremity was defined in the global context of the entire experimental range (i.e. 10–99) was also tested. Findings show similar results but slightly inferior fit compared to the local context model.

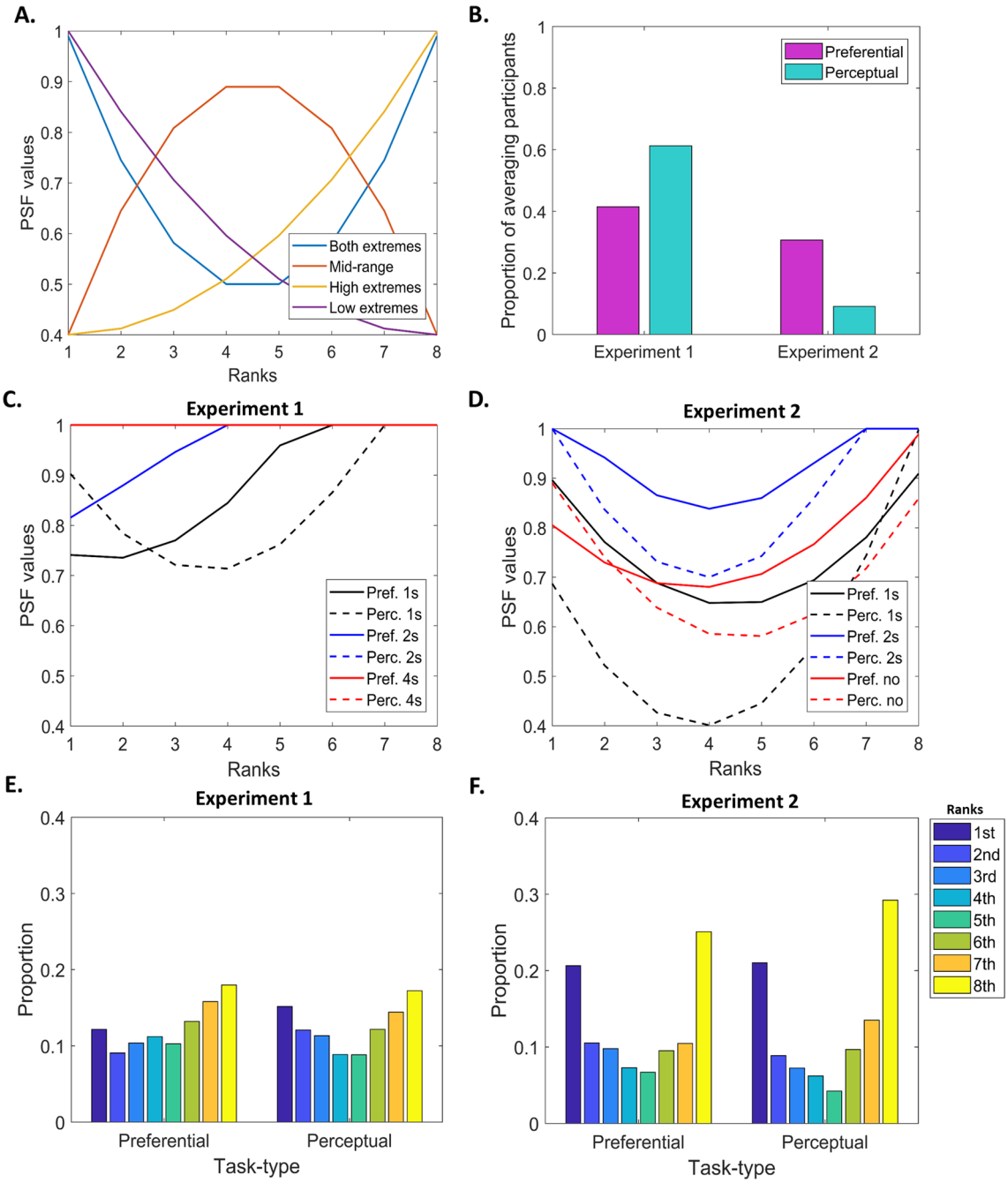

Fig. 3. (A) Example Probabilistic Sampling Functions (PSFs), defined by curvature, symmetry, and the area under the curve. (B) The proportion of participants who were best described by the averaging procedure in Exp. 1 and 2. C & D) The mean values for the best fitting PSFs across participants in Exp. 1 (panel C) and Exp. 2 (panel D). E & F) The proportion of individuals whose maximum PSF value corresponded to each rank (Exp. 1 in panel E and Exp. 2 in panel F). “Pref.” and “Perc.” in the legend indicate the preferential and perceptual tasks respectively, and “1s”, “2s”, "4s" and “no” indicate the time constraints group.

4.2. Integration

Following sampling, the sampled items are combined to calculate the evidence E. The SSIM proposes one of two procedures for integration: averaging, described by Eq. (3) or counting, described by Eq. (4).

$$E = \frac{\sum_j X_j}{N} - R \tag{3}$$

$$E = \Sigma_j(X_j > R) - \Sigma_j(X_j < R) \tag{4}$$

Here j is the sampled items' index, R is the reference value and N is the sample size. In averaging, the mean of the sampled items is calculated and compared to R. In counting, the number of sampled items above R is compared to the number of sampled items below R. Positive values of E correspond to evidence in favor of the array, and negative values of E correspond to evidence in favor of the reference value. To test for significant differences between the decision mechanisms in the preferential and the perceptual tasks, we first compared the performance of the model under two fitting procedure: one where the model was fit simultaneously to both tasks (i.e. estimating a single set of parameters), and one where the model was fit separately to each task. To represent the possibility that people choose their strategies independently for each task, we constructed four versions of the SSIM representing the four possible combinations of strategy and task-type. The first two combinations employ the same integration strategy in both tasks, and we refer to these as *average-average* and *count-count*. The other two combinations posit that people average in the preferential task and count in the perceptual task (*average-count*) or that they count in the preferential task and average in the perceptual task (*count-average*).

To calculate the probability of choosing the array we transform the evidence using the logistic function:

$$p(\text{array}) = \frac{1}{(1 + e^{-[\sigma \times E + \lambda]})} \tag{5}$$

where σ is an additional free parameter that controls sensitivity to the evidence magnitude. The final free parameter, λ determines the tendency to prefer one option over the other regardless of the evidence. $\lambda > 0$ indicates an overall higher probability of choosing the array and $\lambda < 0$ indicates an overall higher probability of choosing the reference value.

We predicted that the sampling strategies, as illustrated by the PSFs, would differ across conditions as a function of task-type and display-time (see details of the OSF pre-registration file in the supplemental materials). Specifically, in Experiment 1, for the preferential task we expected a higher probability of sampling high extremes than low and mid-range items, consistent with the idea that high rewards attract undue attention in risky-choice (e.g. Tsetsos et al., 2012) . By contrast, in the perceptual task we expected to find a higher probability of sampling the mid-range items than the extremes (robust averaging; Vandormael et al., 2017) or equal probabilities among the items, since either can account for the absence of a variance effect in this condition. In Experiment 2 we expected to find large probabilities to sample the high extremes and the low extremes since they were tested in separate trials. We also predicted that the sampling differences in both experiments should be amplified under shorter display durations because there is less opportunity to sample all items and one must therefore sample strategically. Finally, we predicted that in Experiment 1 the *count-average* model would provide the best account since the probabilistic outcome in the preferential task should induce people to count, while the deterministic structure of the perceptual task should have promoted averaging. However, in Experiment 2 we expected the *count-count* model to dominate because both tasks involved probabilistic outcomes that promoted a procedure focused on calculating the probability of "winning".

4.3. Modeling results

For each individual, we first fitted the SSIM to data from both tasks together, implementing either averaging or counting strategy across trials (presented as "average" and "count" models in Table 1). This was followed by separately fitting the data to each version of the SSIM (e.g. for the count-average model we fitted separately counting for the preferential task and averaging for the perceptual task). We used a bounded Nelder-Mead optimization routine (Nelder & Mead, 1965), implemented in Matlab's fminsearch function, to estimate parameter values. Below, we investigate the integration procedures that people used by comparing the performance of each model version at the individual level. We then select the best fitting version for each individual and use the corresponding PSF to examine how selective sampling changed across conditions. The mean values and standard deviations of individuals' best fitting parameters between tasks and experiments are displayed in Table 2[9]. Next, we used individuals' best fitting parameters to simulate choices and produce the model's predictions. This also allowed us to estimate the number of items sampled, i.e. sample size, for each trial. The asterisks in Fig. 2 indicate the predicted mean proportion of choices for the arrays, and show high similarity between model predictions and observed data across all conditions and experiments.

[9] The mean parameters values of α, β and θ do not say much on their own, because the combination of the three forms the unique shape of the PSF that best describes the participant's sampling policy. However, the mean parameters values from the logistic function [σ, λ] should represent the average processing noise and the average bias toward choosing the array across participants.

Table 1
The aggregated AIC scores and weights across participants for each of the model versions in Exp. 1 and 2. Bold indicates the best fitting model in each experiment. The Average-Average model and the Count-Count model posits that people use the same strategies in both the preferential and perceptual task. The Average-Count model posits the use of averaging in the preferential task and counting in the perceptual task; Count-Average posits counting in preferential and averaging in perceptual. The average and count models posit the use of these strategies when data from both tasks was fitted together.

	Experiment 1		Experiment 2	
Model Version	Aggregate	Weight	Aggregate	Weight
Average-average	72974.10	0	63117.27	0
Count-count	73438.85	0	**60828.22**	**1**
Average-count	73864.14	0	61541.70	0
Count-average	**72548.81**	**1**	62403.68	0
Average	75065.12	0	64506.34	0
Count	75244.86	0	62794.61	0

4.3.1. Quantitative comparisons

Table 1 gives the aggregate Akaike information criterion scores (AIC; Akaike, 1973) and the AIC weights (i.e. the conditional probabilities; see Wagenmakers & Farrell, 2004) across participants for each model version. Fig. 3B shows the proportion of individuals best fit by the averaging procedure for each task and experiment. Overall, in Experiment 1 the *count-average* model gave the best fit, indicating that participants were more likely to count in the preferential task and to average in the perceptual task. This was consistent at the individual level (see Fig. 3B), with the majority of participants being best fit by counting in the preferential task (58%) and by averaging in the perceptual task (61%). In Experiment 2 the *count-count* version had the lowest aggregate AIC score, suggesting that probabilistic outcomes promoted counting. This trend was also found at the individual level, with counting providing the best fit for 69% of participants in the preferential task and 91% of participants in the perceptual task. Finally, the model versions that were fitted to each task separately performed better than the model versions that were fitted to both tasks together, therefore supporting our prediction that people employ different choice mechanisms in the preferential and perceptual tasks.

Interestingly, in Experiment 2, counting was more prominent in the perceptual task than in the preferential task. This is consistent with the notion that in the preferential task both averaging and counting can be optimal strategies, depending on whether the EV (average) or the probability to win (count) is maximized. In contrast, under the probabilistic version of the perceptual task, there was no reason for the participants to estimate the arrays' average, so a greater proportion of participants appeared to use a counting strategy.

4.3.2. Sampling strategy

For the following analysis we selected the SSIM version that provided the best fit to each individual. Fig. 3C and D show group-level PSFs for Experiments 1 and 2, respectively, computed as the mean PSF values across individuals. In line with our predictions, under the preferential task in both experiments (solid lines), we found higher probabilities for sampling the high extreme items than mid-range and low extreme items – especially with greater time constraints. In Experiment 1, the PSF values of the perceptual task were often at ceiling (i.e. dashed lines in Fig. 3C), indicating that almost all items were sampled into integration. Although sampling differences between tasks were amplified under greater constraints in Experiment 1, similar PSFs were found across tasks in Experiment 2. Here probabilistic outcomes led participants in both conditions to prioritize the high extremes and the low extremes over the mid-range items (Fig. 3D).

To compare the PSFs at the individual level, we calculated the proportion of PSFs that assigned the largest value to each of the eight item ranks[10]. Consistent with findings at the group level, the majority of participants prioritized the high extremes in both tasks and experiments (see Fig. 3E and F). Overall, 22% of individuals were most likely to sample the highest ranked item, 17% to sample the lowest ranked item, and for the mid-range ranks, the proportions gradually decreased inward - with the lowest proportion of 8% for the 5th ranked item. Interestingly, across tasks and experiments, PSFs that maximized the highest ranked item were more common in conditions that prioritized counting (i.e. preferential in Exp. 1 and perceptual in Exp. 2), possibly because participants were focusing more on counting the number of items above R than below R.

We also simulated the model to estimate the participants' mean sample sizes. That is, when simulating choices according to the best fitting PSFs (to produce the model's predictions), we also simulated the number of items that were sampled from the array on each trial. In both experiments we found evidence for reduced sampling under additional time constraints. In Experiment 1, mean sample size decreased with display-time ($M_{4s} = 6.82$, SD = 0.60; $M_{2s} = 5.98$, SD = 1.27; $M_{1s} = 5.19$, SD = 1.67). In Experiment 2, a similar decrease was found across the 2 s ($M_{2s} = 5.40$, SD = 1.52) and 1 s ($M_{1s} = 4.15$, SD = 1.73) conditions. Curiously, under no time constraints participants sampled less than in the 2 s condition ($M_{no} = 4.88$, SD = 1.87). Taken together with the relatively fast response times observed in this condition, this result suggests that participants also used an incomplete sampling strategy in this condition. Finally, we found that participants sampled fewer items in the preferential task than the perceptual task in Experiment 1

[10] In cases where the PSF showed equal values between the ranks (e.g., probability of 1), the count was divided between the tied ranks.

Table 2
The mean parameters' values (M) and their standard deviation (SD) for the best fitting SSIM across participants, between tasks and experiments.

		α		β		θ		σ		λ	
		M	SD	M	SD	M	SD	M	SD	M	SD
Exp. 1	Preferential	0.14	0.69	0.29	0.39	−0.1	0.7	0.51	0.21	−0.14	0.75
	Perceptual	0.24	0.68	0.43	0.36	0.01	0.67	0.51	0.19	0.05	0.72
Exp. 2	Preferential	0.25	0.62	0.2	0.4	−0.11	0.62	0.49	0.28	−0.01	0.78
	Perceptual	0.39	0.57	0.16	0.46	0.1	0.59	0.62	0.29	−0.06	0.76

($M_{pref.}$ = 5.86, SD = 1.71 vs. $M_{perc.}$ = 6.21, SD = 1.71), with the reverse being true in Experiment 2 ($M_{pref.}$ = 5.00, SD = 2.13 vs. $M_{perc.}$ = 4.62, SD = 2.33).

Finally, the mean parameters values for σ and λ in Table 2 suggest that the different choice patterns and maximization rates we found between the preferential and perceptual tasks were mostly driven by differences in the sampling and integration strategies. Specifically, σ values were similar across tasks in Experiment 1, indicating that differences in processing noise cannot explain the observed differences in maximization rates. Further, the negative mean value for λ in the preferential task of Experiment 1 suggests that the observed risk seeking in this condition was not due to a general bias in favor of gambling. A model recovery test was performed, showing successful recovery of the original PSF shapes, especially at the individual level. (More details of the model recovery are presented in the Appendix A).

5. Probit regression model analysis

Following the work by Vandormael et al. (2017), we also tested the performance of a probit regression model. This model bears some similarity to the SSIM, in that it seeks to understand the relative contribution of each item to an individual's choice. However, it says little about the psychological processes at play. By applying the model to our data, and contrasting its findings with those of the SSIM, we aim to demonstrate the importance of incorporating the notions of strategic and incomplete sampling into models of decision making.

To predict an individual's choice on a given trial the regression model uses the eight items in the array as predictors. Like the SSIM, it rank orders each of the items in an array. A regression weight (i.e. a beta coefficient) is then applied to the value of each item, depending on its rank, and the results are summed and transformed with a probit function to produce a choice probability. In this framework greater beta values indicate a greater impact on choice. Using this analysis Vandormael et al. (2017) found lower beta values for extreme items compared to mid-range items, and inferred that people engaged in robust averaging.

We fitted the regression model separately to each individual and to each task. Mean beta coefficient values are plotted in Fig. 4A. In stark contrast to the findings of the SSIM (Fig. 3), the probit regression analysis showed a robust averaging effect across all groups and conditions, as evidenced by greater beta values for mid-range items than extremes. Thus, without the SSIM's process of strategic and incomplete sampling, the regression model infers that mid-range items had the greatest impact on choice; a conclusion that is difficult to reconcile with the behavioral and modeling results presented above. Comparing the overall fit of the models also yields evidence in favor of the SSIM. To measure the goodness of fit while penalizing for the number of free parameters[11] we calculated the aggregate AIC scores for each participant and model, across tasks and experiments. Our findings showed a better mean fit for the SSIM (aggregate AIC = 132,944; AIC weight = 1) than the probit regression model (aggregate AIC = 139,896; AIC weight = 0). We therefore conclude that the SSIM provides the better account of participants' behavior, both in terms of insight into psychological processes, and in terms of quantitative fit.

In an attempt to reconcile the discrepancy between the SSIM and the probit regression analyses, we tested the probit regression model on a synthetic data set generated by the SSIM (see Appendix A for a full description of the simulation). Since this data set was produced by a process that favored sampling high extremes, the probit model's regression weights should likewise reflect this bias, with higher weights for high extremes. However, as shown in Fig. 4B, larger weights were assigned to the mid-range ranks in both experiments. This mischaracterization suggests that the beta-values do not provide an accurate measure of attention. Instead they reflect the ability of each item to predict choice. Since extreme items were more variable, they were less reliable predictors, and subsequently received lower weighting. This is particularly important for data generated by a counting mechanism (63% of the participants), because this mechanism essentially assigns the same value to all items above (or below) R. Therefore, to fit this behavior, a regression model would need to down-weight the values of the extremes to mimic the 'flattened' value function imposed by counting. To illustrate this notion, we tested the probit model on two synthetic data-sets. One was generated by a pure averaging model, where all items are evenly sampled and averaged to form choice, and the other was generated by a pure counting model, where all items were counted[12]. As exhibited in Fig. 4C, the probit model successfully assigned equal weights to items in the pure

[11] The SSIM had six free parameters, including a dummy parameter that determined whether an averaging or a counting procedure was used. The probit regression model had nine free parameters: eight beta coefficients and an intercept.

[12] In both models, evidence was transformed into choice probability using a logistic function with fixed parameters (i.e. σ = 1, λ = 0).

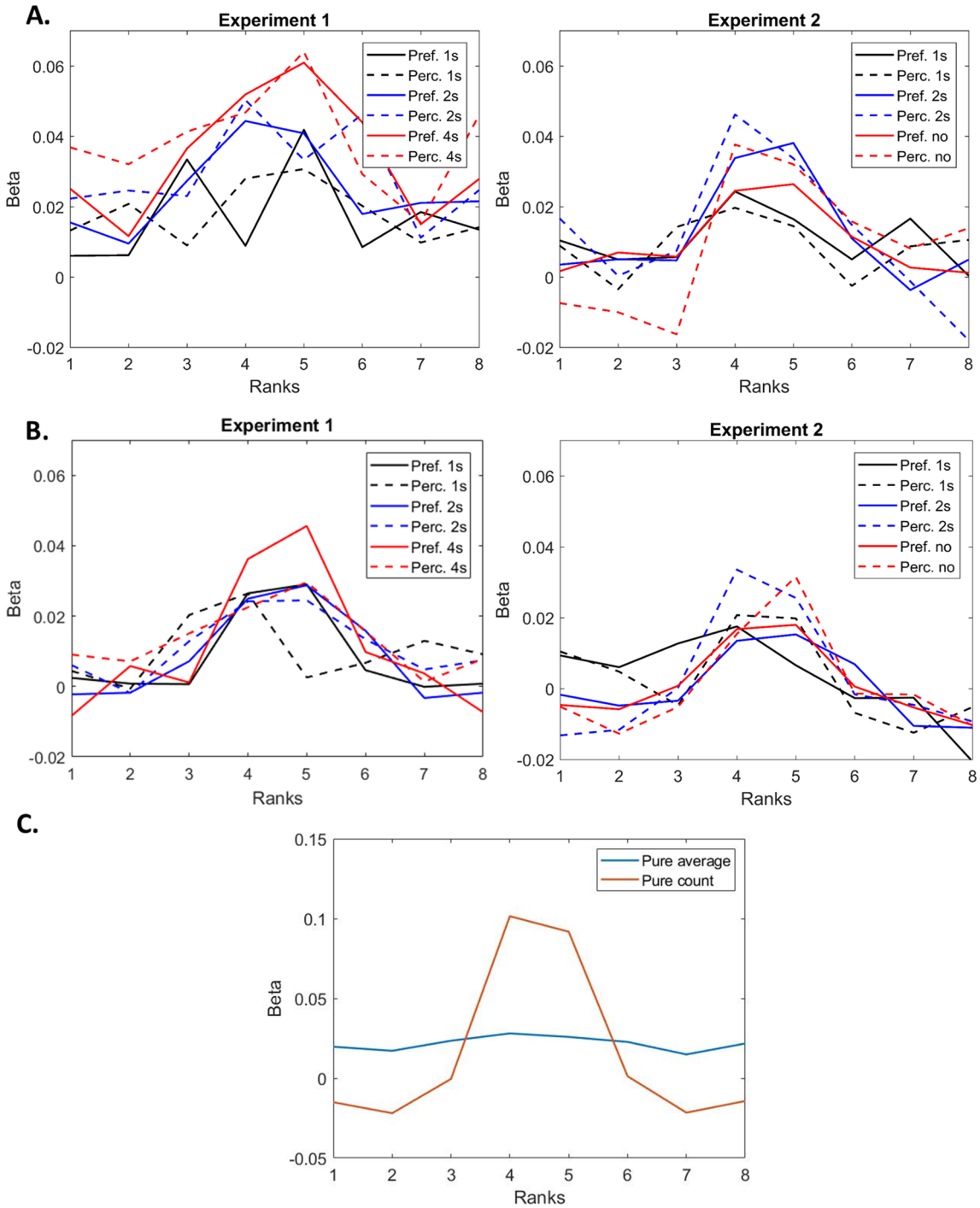

Fig. 4. The mean beta-values from the statistical probit regression model for items' rank, tested separately between task-types, display-time groups and experiments on (A) the original data from Exp. 1 and Exp. 2, (B) a synthetic data-set that was produced by simulating choices according to the participants' best-fit SSIM predictions in Exp. 1 and Exp. 2, and (C) a synthetic data set that was produced by simulating choices according to a pure averaging model and a pure counting model.

averaging model, but assigned larger weights to the mid-range items in the pure counting model. That is, the model erroneously inferred a robust averaging strategy, despite the data being generated from equal weighting of items.

Finally, we also fitted the SSIM to data from Vandormael et al. (2017)'s experiment. Consistent with the authors' findings, results show a clear dominance for averaging over counting (aggregate AIC scores = 45,178.62, AIC weight = 1 vs. aggregate AIC score = 46,293.42, AIC weight = 0). However, similar to our findings from the perceptual task in Exp. 1, the resulting PSFs were often at ceiling (i.e. probability of 1 across ranks), indicating that all items were sampled from the display to form choice. That is, our analysis did not find a preference to sample and integrate the mid-range items. Further, since the SSIM findings implied that participants were likely to sample and average the entire display, we also tested a simple probit regression model with the arrays' average as a single predictor for choice. When comparing the fit and complexity of the three models, the SSIM produced the best fit (aggregate AIC score = 45,158, AIC weight = 1), followed by the simple averaging model (aggregate AIC score = 45,937, AIC weight = 0), and only then by the rank-order probit regression model (aggregate AIC score = 46,006, AIC weight = 0). Thus, suggesting that even between the two statistical models - the rank-order model was redundant.

6. General discussion

In the current study we elucidated an apparent controversy in the literature regarding the impact of extreme items on choice. On the one hand, extreme items seem to have a greater impact on choice in preferential tasks, where participants are incentivized by the outcomes of their choice (Kunar et al., 2017; Lieder et al., 2018; Ludvig et al., 2018; Ludvig et al., 2014; Pleskac et al., 2019; Zeigenfuse et al., 2014). On the other hand, extreme items seem to receive less attention in perceptual tasks that incentivized accurate average estimations of numbers (Li et al., 2017; Vandormael et al., 2017), colors and shapes (De Gardelle & Summerfield, 2011; Michael, de Gardelle, Nevado-Holgado, & Summerfield, 2015), and even faces (Haberman & Whitney, 2010). To explain this differential influence, we posit that choice is goal-driven and guided by attention. As a result, people allocate attention to items in accordance with their task relevance (Tsetsos et al., 2012, 2016; See also Najemnik & Geisler, 2008; Navalpakkam, Koch, Rangel, & Perona, 2010).

To test these assumptions, we developed a behavioral paradigm in which we manipulated task goal and outcome distribution while imposing constraints on attention. However, in contrast to previous studies – which involved presenting items in a fixed sequence (Tsetsos et al., 2012, 2016; Vanunu et al., 2019) – our task involved simultaneous presentation of all items. This allowed participants to sample information freely across the display. To account for our findings and to illustrate how people sampled and processed the information, we introduced the SSIM - a novel cognitive model that assumes people strategically distribute attention across a subset of items in the display, before integrating the sample for choice. Hence, rather than assuming that participants sample and integrate all items – with attention being used to "weight" information – (e.g. Tsetsos et al., 2012, 2016; Glickman, Tsetsos, & Usher, 2018; Vanunu et al., 2019; see also Pleskac et al., 2019; Zeigenfuse et al., 2014), we propose an incomplete sampling procedure that prioritizes task-relevant items, followed by integration via one of two procedures: averaging and counting. In the spirit of bounded rationality (Simon, 1957), the SSIM provides a unified and rigorous account of the flexible process by which individuals tailor their sampling and integration in accordance with task-demands.

In line with our predictions, we found that participants were adaptive (Anderson, 1991; Payne et al., 1988, 1993) and adjusted their sampling and integration strategies in response to task demands. Specifically, in Experiment 1 we found a tendency to sample fewer items in the preferential task and under short display-times, but to focus sampling on the high extremes. This incomplete sampling strategy induced participants to take more risks when variance was high, and in some cases reduced performance (i.e. maximization rates). In Experiment 2, incomplete sampling strategies were found for both tasks, indicating the importance of probabilistic outcomes in encouraging selective sampling. However, in the perceptual task sampling was mostly focused on the high extremes, while in the preferential task it often focused on both extremes. Interestingly, we found differences in sampling even between similar tasks with similar goals. For example, in the preferential task, the lowest extreme ranks were more likely to be sampled in a design with asymmetric distributions (Exp. 2), where the low and high extremes were displayed in separate trials, than in a design with symmetric distributions (Exp. 1).

Using the SSIM, we also found evidence that people selected their integration procedures in response to task goal. In Experiment 1, participants were more likely to use an averaging procedure in the perceptual task – where this was the optimal strategy – and were more likely to use a counting procedure in the preferential task – where probabilistic outcomes might have encouraged participants to consider the probability of "winning". In Experiment 2, both tasks involved probabilistic outcomes, and consequently counting was more prominent in both conditions. Notably, in Experiment 2 averaging was more common in the preferential task than the perceptual task, indicating that some participants in the preferential condition based their decisions on the gambles' EVs. This result emphasizes the importance of considering both strategies for preferential choice.

In contrast to the findings of Vandormael et al. (2017), analysing our results with the SSIM did not reveal robust averaging. On the contrary, in both experiments we found that the majority of participants were most likely to sample the high extreme items. These findings are consistent with previous studies that have found similar behavioral and computational results across preferential and perceptual choices (Hotaling, Cohen, Shiffrin, & Busemeyer, 2015; Tsetsos et al., 2012, 2016). One potential explanation for why people would focus on the high extremes in the perceptual task – where accuracy is the goal – is that task goal is not the only factor influencing behavior. Perhaps more bottom-up processes, such as salience (see Itti & Koch, 2000), attract attention to extreme items (Krajbich, Armel, & Rangel, 2010; Kunar et al., 2017), independent of task demands. For example, due to the Gaussian distribution of outcomes, participants in our study saw values between 50 and 60 on nearly every trial, but only encountered values above 90 or below 20 on approximately 2% of trials. This difference may have led to a "pop-out" effect for rare values that attracted participants' attention in task-irrelevant ways (Wang, Cavanagh, & Green, 1994).

In contrast, when analyzing our data using only a probit regression model, we found a pattern similar to that of Vandormael et al. (2017). Here, the greater impact of mid-range items on choice is indicated by higher beta values across all conditions and experiments. This apparent contradiction points to fundamental differences between the SSIM and probit models. The most obvious is that the SSIM describes a process of incomplete information sampling, while the probit model assumes all information is used, with regression weights representing the predictive power of each item rank. Our findings suggest that these weights may not reflect the allocation of attentional resources across items. Instead, they may simply indicate features of the environment. For instance, mid-range items may have served as better predictors for choice because their values were more consistent across trials, while the values of the extreme items varied considerably between trials and variance conditions. Heavily weighting these extremes would produce erratic behavior in the probit model, which would explain why the model employs robust averaging. Future studies could investigate this further using different outcome distributions (e.g. uniform or U-shaped) to produce different regression weights.

Another major difference between the SSIM and the probit regression model lies in the integration process. While the probit regression model uses a weighted summation of all items (i.e. similar to averaging), the SSIM suggests that participants might integrate items differently based on task demands. In fact, across all experiments and conditions we found that the majority of participants (63%) were better described by counting than averaging. Critically, a counting mechanism assigns all items above (or below) R the same value. Therefore, to fit behavior produced by this process, a regression model would need to down-weight the magnitudes of extreme items to equate their impact with that of mid-range items. This highlights the importance of considering different integration methods when characterizing the mechanisms underlying judgment and choice[13] (see Section 5).

In summary, we contend that the SSIM provides a superior quantitative fit to participants' choice data, indicating a more accurate representation of behavioral patterns than that provided by the probit regression. Moreover, it yields an arguably more psychologically plausible account of how people collect and process information. Applying the model to other paradigms requiring selective-attention may yield similar insights. At present, we use a one-dimensional PSF – based only on rank order – but new studies can investigate other dimensions relevant to people's processing, e.g. magnitude, familiarity, location, temporal order, color, and orientation. In this way, the SSIM represents a valuable tool for analyzing and understanding how one's strategic goals, task-demands, and environmental constraints influence decisions in several domains.

One issue that remains unclear is how people direct their attention to items in the array that they have yet to see. One possible explanation is that a low-level visual process scans everything into the perceptual system (Itti & Koch, 2000; Koch & Ullman, 1985; Niebur & Koch, 1996), at which point a higher-level 'decisional' process selectively samples items for integration (Kanan, Tong, Zhang, & Cottrell, 2009; Underwood, Foulsham, van Loon, Humphreys, & Bloyce, 2006; see also Hastie & Park, 1986). Together, the PSF describes the end results of the bottom-up and top-down processes that governs attention. Future studies should try to disentangle these processes by testing the impact of bottom-up features, like saliency, on choice while using different scales for the PSF to determine which dimension - saliency or rank - is more prominent in sampling.

7. Conclusion

In the current study, we found that people were adaptive and adjusted their sampling and integration strategies according to task demands. In both experiments they employed incomplete sampling strategies that prioritized the information relevant to their goal. They likewise used integration procedures that matched task-demands and goal. Future work is needed to further understand how people strike a balance between maximizing accuracy (or earnings) and conserving time and cognitive resources.

Author note

This research was supported by the Australian Government International Research Training Program (IRTP) scholarship to Y. Vanunu and an Australian Research Council Discovery Project to B. R. Newell (DP160101186). We thank Chris Donkin and Emmanouil Konstantinidis for useful discussions and feedback on this project.

Appendix A

A1. Model recovery

To test the reliability of the SSIM, we generated a synthetic data set according to participants' best fitting parameters, and fitted the SSIM to this data to test model recovery. Since our primary interests lie with sampling behavior, we examined whether the generating PSF shapes, rather than parameter values, were recovered. As shown in Fig. A1, all of the key results of the model analysis were well recovered. 95% of the participants had the same best fitting model version as in the original analysis. Further, the shape of the PSFs were also recovered well, showing the same patterns of oversampling the extremes - especially the high extremes - across tasks and experiments.

[13] Vandormael et al. (2017) tested the prediction power of counting against averaging in a competitive regression analysis. Although both procedures accounted for a significant component of the variance, the authors rejected the counting model on the grounds that its regression coefficient was smaller than that for averaging.

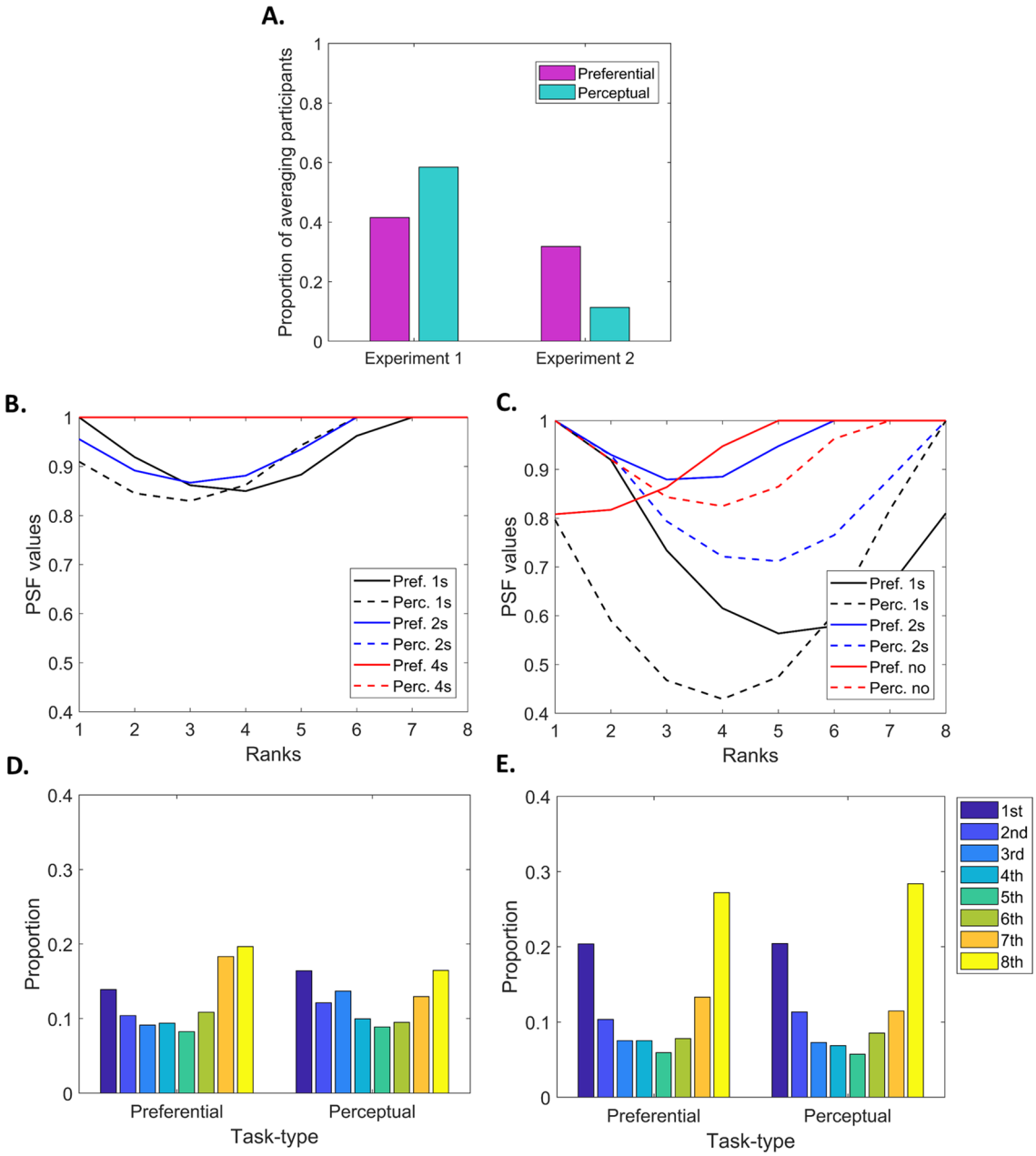

Fig. A1. Findings of the recovery analysis including: (A) The proportion of participants who were best described by the averaging procedure in Exp. 1 and 2. (B & C) The mean values for the best fitting PSFs across participants in Exp. 1 (panel C) and Exp. 2 (panel D). (D & E) The proportion of individuals whose maximum PSF value corresponded to each rank (Exp. 1 in panel E and Exp. 2 in panel F). "Pref." and "Perc." in the legend indicate the preferential and perceptual tasks respectively, and "1s", "2s" and "no" indicate the time constraints group.

Appendix B. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cogpsych.2020.101274.

References

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In B. N. Petrov, & F. Csaki (Eds.). *Second international symposium on information theory* (pp. 267–281). Budapest: Akademiai Kiado.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98*(3), 409–429.

Ariely, D., & Zakay, D. (2001). A timely account of the role of duration in decision aking. *Acta Psychologica, 108*(2), 187–207.
Brusovansky, M., Vanunu, Y., & Usher, M. (2019). Why we should quit while we're ahead: When do averages matter more than sums? *Decision, 6*(1), 1–16.
De Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences, 108*(32), 13341–13346.
Doane, D. P., & Seward, L. E. (2011). Measuring skewness: A forgotten statistic? *Journal of Statistics Education, 19*(2).
Dutilh, G., & Rieskamp, J. (2016). Comparing perceptual and preferential decision making. *Psychonomic Bulletin & Review, 23*(3), 723–737.
Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., ... Heiberger, R. (2012). *Package 'car'*. Vienna: R Foundation for Statistical Computing.
Gigerenzer, G. (2004). Fast and frugal heuristics: The tools of bounded rationality. In D. J. Koehler, & N. Harvey (Eds.). *Blackwell handbook of judgment and decision making* (pp. 62–88). Malden: Blackwell Publishing.
Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62*, 451–482.
Gigerenzer, G., & Goldstein, D. G. (1999). Betting on one good reason: The take the best heuristic. In G. Gigerenzer, P. M. Todd, & The ABC Research Group (Eds.). *Evolution and cognition. Simple heuristics that make us smart* (pp. 75–95). New York, NY, US: Oxford University Press.
Glickman, M., Tsetsos, K., & Usher, M. (2018). Attentional selection mediates framing and risk-bias effects. *Psychological Science, 29*(12), 2010–2019.
Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention, Perception, & Psychophysics, 72*(7), 1825–1838.
Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review, 93*(3), 258–268.
Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description–experience gap in risky choice: The role of sample size and experienced probabilities. *Journal of Behavioral Decision Making, 21*(5), 493–518.
Hertwig, R., & Pleskac, T. J. (2008). The game of life: How small samples render choice simpler. In N. Chater, & M. Oaksford (Eds.). *The probabilistic mind: Prospects for rational models of cognition* (pp. 209–236). Oxford University Press.
Hertwig, R., & Pleskac, T. J. (2010). Decisions from experience: Why small samples? *Cognition, 115*(2), 225–237.
Hotaling, J. M., Cohen, A. L., Shiffrin, R. M., & Busemeyer, J. R. (2015). The dilution effect and information integration in perceptual decision making. *PloS One, 10*(9), e0138481.
Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10–12), 1489–1506.
Kanan, C., Tong, M. H., Zhang, L., & Cottrell, G. W. (2009). SUN: Top-down saliency using natural statistics. *Visual Cognition, 17*(6–7), 979–1003.
Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology, 4*, 219–227.
Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience, 13*(10), 1292–1298.
Kunar, M. A., Watson, D. G., Tsetsos, K., & Chater, N. (2017). The influence of attention on value integration. *Attention, Perception, & Psychophysics, 79*(6), 1615–1627.
Li, V., Castañón, S. H., Solomon, J. A., Vandormael, H., & Summerfield, C. (2017). Robust averaging protects decisions from noise in neural computations. *PLoS Computational Biology, 13*(8), e1005723.
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.
Ludvig, E. A., Madan, C. R., McMillan, N., Xu, Y., & Spetch, M. L. (2018). Living near the edge: How extreme outcomes and their neighbors drive risky choice. *Journal of Experimental Psychology: General, 147*(12), 1905–1918.
Ludvig, E. A., Madan, C. R., & Spetch, M. L. (2014). Extreme outcomes sway risky decisions from experience. *Journal of Behavioral Decision Making, 27*(2), 146–156.
Ludvig, E. A., & Spetch, M. L. (2011). Of black swans and tossed coins: Is the description-experience gap in risky choice limited to rare events? *PloS One, 6*(6), e20262.
Madan, C. R., Ludvig, E. A., & Spetch, M. L. (2014). Remembering the best and worst of times: Memories for extreme outcomes bias risky decisions. *Psychonomic Bulletin & Review, 21*(3), 629–636.
Madan, C. R., Spetch, M. L., & Ludvig, E. A. (2015). Rapid makes risky: Time pressure increases risk seeking in decisions from experience. *Journal of Cognitive Psychology, 27*(8), 921–928.
Michael, E., de Gardelle, V., Nevado-Holgado, A., & Summerfield, C. (2015). Unreliable evidence: 2 sources of uncertainty during perceptual choice. *Cereb Cortex, 25*(4), 937–947.
Najemnik, J., & Geisler, W. S. (2008). Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision, 8*(3) 4-4.
Navalpakkam, V., Koch, C., Rangel, A., & Perona, P. (2010). Optimal reward harvesting in complex perceptual environments. *Proceedings of the National Academy of Sciences, 107*(11), 5232–5237.
Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal, 7*(4), 308–313.
Niebur, E., & Koch, C. (1996). Control of selective visual attention: Modeling the 'where' pathway. In D. Touretzky, M. Mozer, & M. Hasselmo (Eds.). *Neural Information Processing Systems (NIPS 8)* (pp. 802–808). Cambridge, MA: MIT Press.
Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 534–552.
Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker.* Cambridge, England: Cambridge University Press.
Pleskac, T. J., Yu, S., Hopwood, C., & Liu, T. (2019). Mechanisms of deliberation during preferential choice: Perspectives from computational modeling and individual differences. *Decision, 6*(1), 77–107.
Simon, H. A. (1957). *Models of man: Social and rational.* Wiley.
Spitzer, B., Waschke, L., & Summerfield, C. (2017). Selective overweighting of larger magnitudes during noisy numerical comparison. *Nature Human Behaviour, 1*(8), 0145.
Thorngate, W. (1980). Efficient decision heuristics. *Behavioral Science, 25*(3), 219–225.
Tsetsos, K., Chater, N., & Usher, M. (2012). Salience driven value integration explains decision biases and preference reversal. *Proceedings of the National Academy of Sciences, 109*(24), 9659–9664.
Tsetsos, K., Moran, R., Moreland, J., Chater, N., Usher, M., & Summerfield, C. (2016). Economic irrationality is optimal during noisy decision making. *Proceedings of the National Academy of Sciences, 113*(11), 3102–3107.
Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review, 79*(4), 281–299.
Underwood, G., Foulsham, T., van Loon, E., Humphreys, L., & Bloyce, J. (2006). Eye movements during scene inspection: A test of the saliency map hypothesis. *European Journal of Cognitive Psychology, 18*(03), 321–342.
Vandormael, H., Castañón, S. H., Balaguer, J., Li, V., & Summerfield, C. (2017). Robust sampling of decision information during perceptual choice. *Proceedings of the National Academy of Sciences, 114*(10), 2771–2776.
Vanunu, Y., Pachur, T., & Usher, M. (2019). Constructing preference from sequential samples: The impact of evaluation format on risk attitudes. *Decision, 6*(3), 223–236.
Wagenmakers, E. J., & Farrell, S. (2004). *Psychonomic Bulletin & Review. 11*(1), 192–196.
Wang, Q., Cavanagh, P., & Green, M. (1994). Familiarity and pop-out in visual search. *Perception & Psychophysics, 56*(5), 495–500.
Zeigenfuse, M. D., Pleskac, T. J., & Liu, T. (2014). Rapid decisions from experience. *Cognition, 131*(2), 181–194.