Noisy Integration of Value Differences

Silvio Ravaioli

August 28, 2019

Abstract

Many everyday decisions require the evaluation and comparison of alternatives across multiple dimensions. Recent experimental evidence suggests that humans suffer from systematic biases even in simple averaging tasks (Spitzer, Waschke, and Summerfield 2017, Li et al. 2018). Random utility models (in economics) and noisy perception models (in psychophysics) usually assume that the evaluation of each observation is implemented in isolation, and this assumption is at odds with extensive field and experimental evidence of context effect. In order to study human information integration, we use an averaging task with binary choice between compound lotteries. Participants observe two sequences of simple lotteries, whose winning probabilities are depicted using bars with different heights, and choose which compound lottery to implement. After presenting the main features of the choice data, we show that models that assume integration of differences fit the data better than their counterparts with separate evaluation of individual values. Finally, we compare the behavior in two treatments with different processes generating the trials values. The same distribution of value differences is generated by two distinct single-value distributions. The two treatments show significant differences, suggesting that the way differences are generated is also relevant. This result is consistent with a model of salience a' la Bordalo, Gennaioli, and Shleifer (2012), with focusing weight in favor of observations that are unusual or surprising relative to the reference. We discuss implications of noisy integration of value differences in a buyer-seller obfuscation pricing setting, and as a possible explanation for context effect and violation of stochastic transitivity.

Keywords: Multidimensional choice, Context effect, Noisy value representation, Salience

1 Introduction

Many everyday decisions require the evaluation and comparison of alternatives that differ across multiple dimensions. Cereals in the supermarket may differ in price, nutrient content, and package size, banks offer contracts with different fees, insurance companies create plans with various premium and coverage, and political parties advance platforms that diverge on many economic and social issues. On the one hand, having multiple options allows to improve the final choice, as agents also differs in their preference and needs. On the other hand, we commonly experience difficulty in comparing and choosing among several options, especially if they are widely different.

In some of the aforementioned choice problems the value of each option can be reduced from many dimensions to a single numerical values, that represents its value. Consider an agent that needs to do grocery shopping in one store among the only two available ones in the neighborhood. The agent has recently visited both stores and has observed all the prices, and now she wants to purchase the products itemized in a shopping list. We assume that the shopping list is fixed, therefore the agent's purchases are not sensitive to prices. All the items are available in both stores, and her goal is to minimize the total cost for completing the grocery. If the agent has perfect memory, she will calculate the total cost of purchasing the items in each store, and pick the cheapest one. But it is plausible to include some imprecision in the recall of prices from memory. If every item in the first store is systematically cheaper than the corresponding one in the second store, the final choice will be easy even with poor memory. This is usually not the case, and one store may have cheaper produce but more expensive meat, making the comparison difficult and increasing the chance of eventually picking the overall expensive store.

We are interested in understanding how a decision maker (DM) evaluates the vectors of dimensions and eventually selects one option under dimension-by-dimension comparison. Such a way to compare alternatives is natural and further facilitated by many online stores: Amazon.com, for example, displays for each product a table labeled "Compare with similar items" that allows to note the dissimilarities of the selected product with respect to three (not always) *similar* alternatives. Understanding better how comparison plays a role in the decision process gives a twofold motivation for this paper. From a theoretical perspective, our results provide novel insights for the design of a buyer-seller price obfuscation model where the sophisticated firm can take advantage of the biases of a naive buyer. The recent literature about multidimensional pricing strategy adopts similar frameworks, with various types of biases affecting the consumer side. From an empirical perspective, our result speak directly to the discrete choice models that are often used to estimate demand. Commonly used Random Utility Models as-

sumes that the valuation of separate elements occurs in isolation, whereas our results suggest that explicit comparison is an important feature of human decision process, and neglecting it may lead to severe biases in the estimation.

Random Utility Models (RUM) are often used in the economic literature to introduce heterogeneity across decision makers in cross-sectional data, or across time in panel data. Starting from the early work of Thurstone (1927), Block and Marschak (1960), and Luce (1959), this family of models have been widely used to capture discrete choices. One subtle assumption that is typically used is that each option is evaluated in isolation, and the associated random shock is independent from the other elements in the choice set. This feature is at odd with the vast empirical and experimental literature about context effect, that highlights the non-neutrality of the elements in the choice set, with particular emphasis for the role of "decoys" of various nature.

In this paper we use the experimental methodology to test and discuss the aforementioned assumption of evaluation of attributes in isolation. We use a simple perceptual *averaging* task to create a controlled environment in which both experimenter and subject are aware of all the parameters of interest. In each trial, the decision maker observes two options, each characterized by a vector of values (dimensions), and is asked to choose the option associated with the highest mean value.

The data collected in the perceptual task display various systematic patterns at odds with traditional RUM, but consistent with a generalized model that allows direct comparison between options. First, choice is stochastic, and the probability of choosing an option depends on its overall distance from the value of the other alternative. Second, accuracy depends on the *similarity* between two options across dimensions. Intuitively, two options are similar when they display small (absolute) differences along each dimension. This result suggests that the DM encodes a noisy representation of the difference, instead of two values in isolation. A model fitting exercise suggests that the variance in the encoding process is monotonic in the difference, in a fashion similar to a multiplicative error. Finally, the encoding of differences is not independent from the data generating process. We observe context effect consistent with salience models, where the DM focuses on unusual, odd values.

The rest of the paper is structured as follows. Section 2 develops the connection of this paper with the existing literature. Section 3 introduces the experimental design. Sections 4 and 5 contain a series of reduced-form results and systematic choice patterns in the experimental task, followed by the description and fitting of various models that allow for *noisy perception* of

the true values. Section 6 concludes with a discussion of the implications of our results, with particular emphasis on context effect and violation of stochastic transitivity.

2 Related Literature

This paper shares part of its motivation with three main research topics: multidimensional pricing strategies, noisy integration of decision information, and context effect in consumers' choice.

The multidimensional pricing literature includes a family of models where sophisticated sellers exploit naive buyers' biases (Ellison 2005, Spiegler 2006, Brown, Hossain, and Morgan 2010, Piccione and Spiegler 2012]. Buyers that are characterized by bounded rationality may ignore or forget relevant information, or adopt consideration sets for their choices, and the seller strategically design prices in order to maximize own profits. On a similar note, Gabaix and Laibson (2006) focus on the specific case of shrouded attributes, where part of the product's characteristics are omitted by choice of the seller, and the buyer fails in extrapolating the missing information. These papers typically assume a specific form of bounded rationality on the buyer side, without the need of providing compelling argument for it, and analyze the strategic problem for the seller. Our paper takes a different approach, as we focus only on the buyer's side of the problem in order to provide a more grounded theory behind these examples of bounded rationality. We aim to provide a fundamental piece to be included in strategic models, and we briefly discuss in Section 6 two possible applications of our results.

Extensive empirical and experimental evidence suggest that systematic biases in buyers' behavior actually exist. Consumers may select a dominated option (as shown by Sinaiko and Hirthb 2011 in the case of healthcare plans) and fail in integrating of relevant information (Chetty, Looney, and Kroft 2009 show that consumers underreact to taxes that are not salient). Additional insights come from the experimental literature, where the choice setting can be fully observed. Consistent departures from the unbiased behavior can be observed even in simple averaging tasks, where the participants are asked to choose the alternative associated with the higher average value, as in Tsetsos et al. (2016), providing evidence of noisy integration of decision information and "selective gating" in favor of the best local option. In simple problems where dimensions are equally relevant, an optimal observer integrates them with the same weight. Humans deviate systematically from this prediction by integrating information unequally based on the distance from the expectation. Depending on the setting, either overweight of extreme values (Spitzer, Waschke, and Summerfield 2017) or robust averaging (Li et al. 2018) are observed. In a multidimensional pricing model, overweight [robust averaging] would be interpreted

as higher [lower] price elasticity in the dimensions whose price is far from the average. Both these patterns can be shown to be optimal based on the nature of the *noise* that occurs in the decision process. Overweight is consistent with *early* noise associated with imprecise perception, whereas *late* noise is consistent with imperfect memory and accommodates for robust averaging. Our analysis suggest an alternative explanation to late noise in the encoding of information in isolation, and to selective gating in the direct comparison between attributes.

Despite the overwhelming evidence of biased and context-dependent encoding of information, we still do not have a unified theory that is able to capture all the relevant features. A vast and discordant literature discusses context effect in consumers' choice both in theoretical and experimental settings. Various models have different or even opposites predictions [Tversky and Simonson (1993), Kivetz et al. (2004), Soltani, de Martino and Camerer (2012), Bordalo, Gennaioli and Shleifer (2013), Koszegi and Szeidl (2013), Bushong et al. (2017), Natenzon (2018), Landry and Webb (2019)]. Stochastic transitivity violations, that emerges under context effect, would not occur if information was encoded in isolation for each item. In Section 6 we discuss how stochastic transitivity violations in binary choices can appear in our setting and are consistent with the models we introduce in Section 5.

3 Experimental Design

3.1 Binary averaging problem

The setting of the experiment can be summarized by the following binary averaging and choice problem. Consider two options, labeled by $i \in \{L, R\}$. Each option is described by a vector of T numerical values $X^i \equiv \{x_t^i\}_{t=1}^T$ and all the values are generated from the same distribution, independently across options and dimensions $x_t^i \sim F(\cdot)\ \forall i, t$. The *true* value of an option is linear in the dimensions' values, and is the average of its T dimensions:

$$v(X^i) = \frac{1}{T}\sum_{t=1}^{T} x_t^i \tag{1}$$

The DM compares the two options one dimension at the time by observing the sequence of pairs $\{(x_t^L, x_t^R)\}_t$, and at the end she chooses one of the options. Her payoff is equal to the true value of the selected option.

We allow the value of each option *perceived* by the DM to be different from the true one. Possible reasons for this inconsistency include measurement error, perceptual noise, and imper-

fect memory. The DM will implement her choice based on the perceived values, but the payoff she receives is the true value of the chosen option. We assume that the DM wants to maximize own payoff, and if no other friction is introduced this is equivalent to maximize the probability of choosing the option with the highest value.

3.2 Averaging task

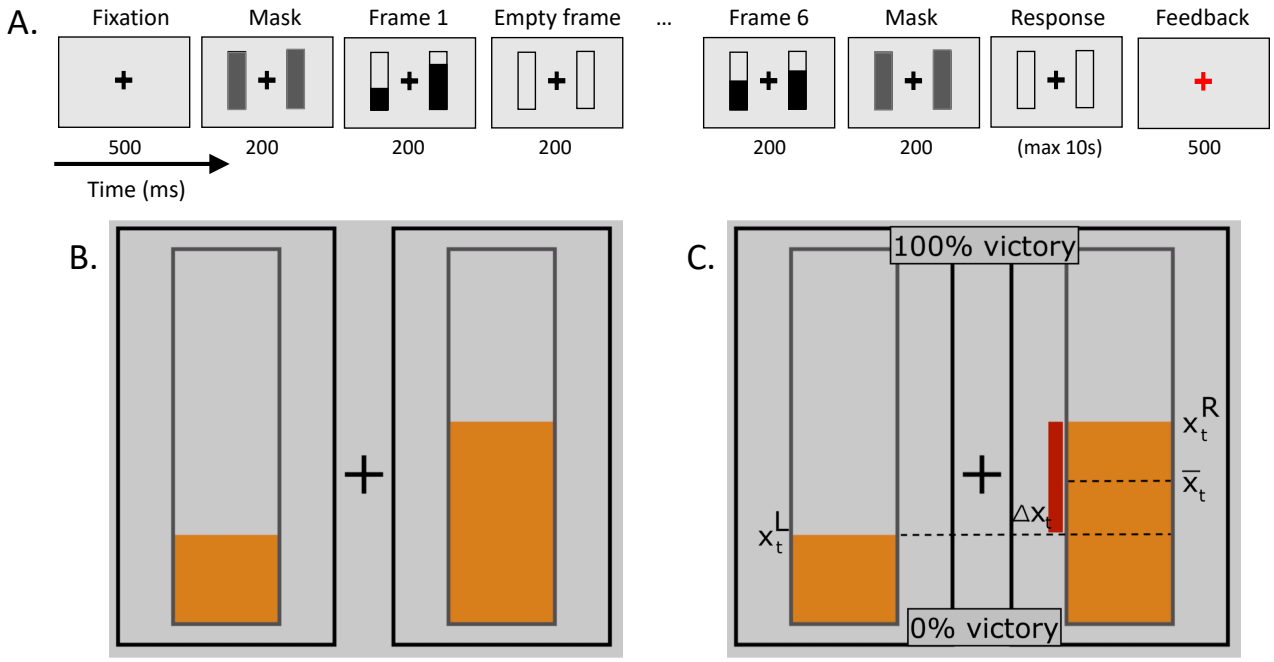


Figure 1: Behavioral task. (A) Trial schematic: Participants view two streams of bars representing compound lotteries, choose one, and receive immediate feedback about the outcome (color of the fixation cross). (B) Frame: colors and proportion of a frame. Orange bars' heights indicate the probability of winning one point in the simple lottery. (C) Main elements of the frame: x_t^i indicate the percentage probability of winning one point in option $i \in \{L, R\}$ and frame $t \in \{1, 2, ..., 6\}$. Our notation include $\Delta x_t \equiv x_t^L - x_t^R$ and $\bar{x}_t \equiv \frac{x_t^L + x_t^R}{2}$.

Our experimental design follows closely the averaging paradigm used by Tsetsos et al. (2016). In each trial, a sequence of six frames is presented. Each frame contains a pair of bars (one on the left and one on the right of the screen). The participant observes all the frames, then she chooses one option and she receives a reward that depends on the value of the chosen option. We adapt Tsetsos' design to address some common experimental concerns. First, all the values (bars' lengths) are generated by the same data generating process, and the participants are well informed about that (the Instructions are available in Appendix C). Second, we adopt a reward structure that relies on the construction of compound lotteries. The height of the bar represents

the probability of winning one point in a simple lottery. An option represents a compound lottery, with uniform probability over the six simple lotteries that are associated with it.

3.3 Treatments and stimuli

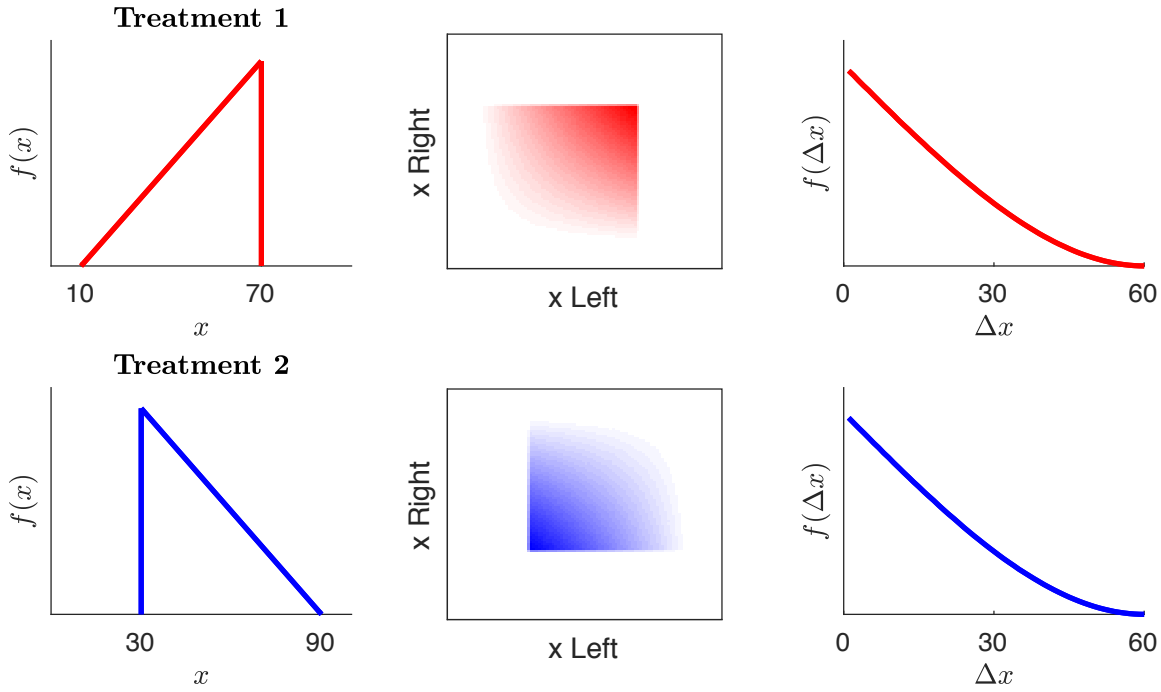


Figure 2: Value distributions used to generate data in the two treatments. Top/Bottom rows indicate the treatments. Left: PDF of the single stimulus (i.e. winning probability in the simple lottery, displayed as bar height). Center: Joint PDF of the pair of independent stimuli in a frame. Right: PDF of the absolute difference $|\Delta x_t|$ between higher and lower value in a frame.

Participants are randomly assigned to one out of two treatments. The experiments are identical across treatments, with a simple difference about the data generating process adopted. Figure 2 shows the PDF of each treatment: in both cases we use a triangular distribution with a range of 60 percentage points, and mean equal to 50%. The two treatments are associated with different supports (10-70 or 30-90 percentage points), as well as different PDFs (upward or triangular distributions). By doing so we obtain two environment with different marginal and joint distribution of values, but the same distribution of differences within frames. In particular, most of the frames will have small differences, with a minimum of 0 and a maximum of 60 percentage points. The comparison of these two treatments will be useful, as a model that takes into account only value differences should predict the same choice pattern in both environments.

3.4 Procedure

The experiment involves 300 incentivized trials, each comprised of six frames as the one showed in Figure 1B. The fixation cross is displayed on the screen during the whole trial, and it changes color after the choice to provide an immediate and intuitive feedback. The cross becomes green or red based on the positive or negative outcome of the lottery. No feedback is provided about the chosen option being the one with the highest average.

At the beginning of the experiment, participants navigate freely the instructions and are able to ask clarifications to the experimenters if necessary. Detailed experimental instructions are provided in Appendix C and include the description of the task and the reward structure [with four unrewarded and *simplified* practice trials], information about the data generating process [using plain English, graphical representation of the frequency of stimuli, and free practice], and five unrewarded practice trials.

The trials are grouped in 4 blocks of 75 trials each, and participants are encouraged to take a small break between blocks in order to minimize discomfort. Instructions and experiment took on average 40 min.

3.5 Apparatus, participants, and payment

The experiment was run in CELSS (Columbia Experimental Laboratory of Social Sciences, Columbia University, New York, USA) between February and March 2019. The participants responded to the task demands using an ordinary keyboard, and navigated the instruction pages using a mouse. Experiments and stimuli were coded in MATLAB (Release 2018b) using Psychotoolbox 3 (Psychophysics Toolbox Version 3).

Fifty-one healthy volunteers were recruited using the platform ORSEE (Online Recruitment System for Economic Experiments) and were naive to the main purpose of the study. All participants provided written, informed consent. 26 subjects completed treatment 1 and 25 subjects completed treatment 2. The whole experimental session comprised of the main task described above and two ancillary tasks (not included in this paper), for a total of 800 trials (about 80 minutes in the laboratory, including instructions and payment). Participants were paid a variable amount of money according to task performance: they received 20 cents of dollars for every point collected (across 800 trials) after the first 300. This means that 400 points (the average score of a random DM) would pay \$20, 401 points would pay \$20.20, and so on. Participants were informed of the payment structure and distribution of scores based on their accuracy, and were guaranteed a minimum payment of \$10. On completion of the experiment, the subjects received payment in cash. Final payments vary based on participants' skill and luck, and ranged from \$19.40 to \$28.20, with an average of \$24.60.

4 Reduced-form Results

In this section we present the main descriptive statistics of the data. Task difficulty was calibrated based on the performance of the participants in a pilot study conducted in November 2018. In our experiment we observe an overall 74% accuracy. This score indicates the fraction of times participants chose the option with the highest overall value, averaging across participants. Participants have significantly different accuracy, ranging from 62% to 84%, but the populations of subjects assigned to the two treatments are not significantly different. Appendix A contains robustness checks that address usual concerns about data quality. We do not observe learning (i.e. performance improvement over time), consistently with the extensive instructions and practice before the task.

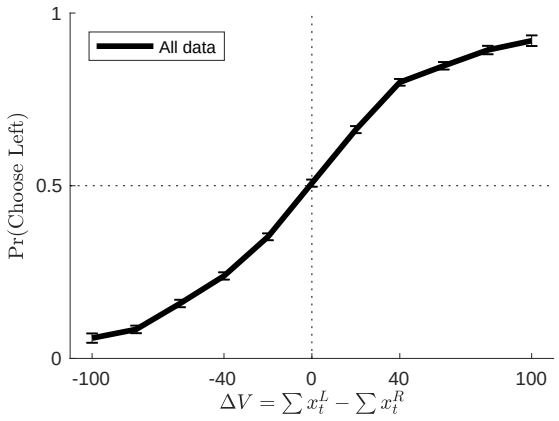


Figure 3: The probability of choosing option L, plotted as a function of the overall difference in values between options L and R. Standard errors are included.

Figure 3 shows the stochastic relation between choices and average value displayed. We define $\Delta V \equiv \Sigma_{t=1}^{6}(x_t^L - x_t^R)$ as the total differences between the sum of values of option L and R. The probability of choosing option L is a continuously increasing function, rather than a step function. This pattern is consistent with extensive evidence in perceptual judgment tasks (starting from the work of Weber 1978 [originally 1834]) as well as tasks of choice under uncertainty (Mosteller and Nogee 1951, Khaw, Li, and Woodford 2017).

Additionally to the monotonic stochastic pattern that emerge by comparing trials with different ΔV, choices are systematically affected by other features of the observed values. Figure 4 captures two peculiar patterns.

Each trial can be characterized not only by the *overall* difference ΔV, but also by the frequency of local wins $\Delta W \equiv \Sigma_{t=1}^{6}\mathbb{1}(x_t^L > x_t^R)$, and by the *average* frame difference $\Delta S \equiv \frac{1}{6}\Sigma_{t=1}^{6}|x_t^L - x_t^R|$, that provides an intuitive measure of similarity between the options. ΔW counts in how many

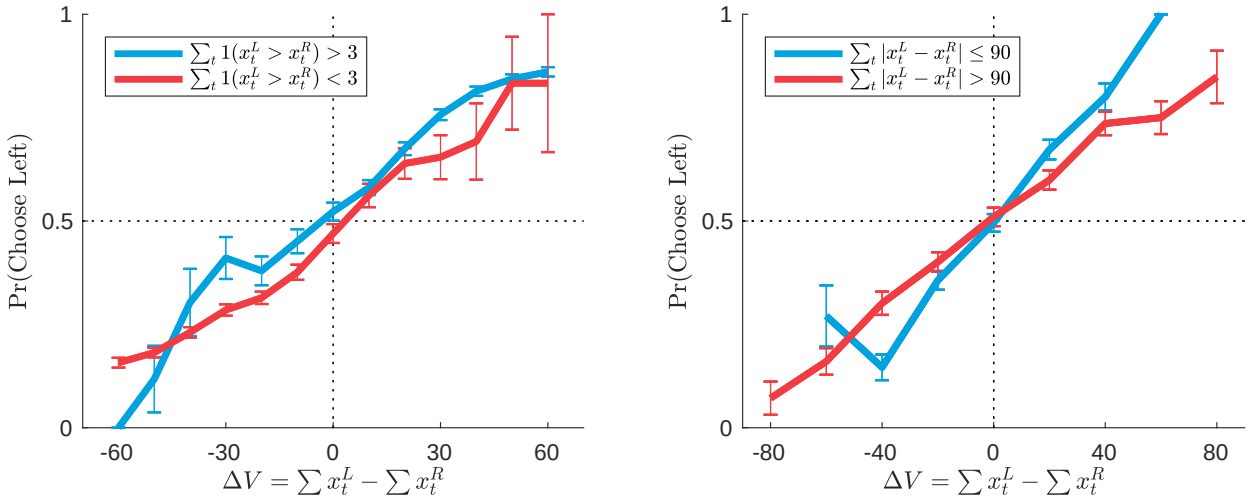


Figure 4: Probability of choosing option L after controlling for number of local wins ΔW and similarity ΔS. Left: Trials are grouped by number and direction of local wins, as described in text. Right: Only trials with no dominating local winner. Trials are grouped by similarity ΔS.

pairwise comparisons the value x_t^L is higher than x_t^R. Figure 4.a displays the different choice probabilities in the trials in which L is the dominating local winner (i.e. L is higher in 4 or more frames, in blue) or R is the dominating local winner (in red). Trials without a dominating local winner are omitted. We observe a systematic bias in favor of the dominating local winner: option L is chosen significantly more often when L is winning in more frames, after controlling for the overall difference ΔV. This pattern would be consistent with the simple heuristic of counting the number or frames in which an option is the winning one, ignoring the magnitude of the difference, and choosing the dominating local winner.

After controlling for the *dominating local winner*, another feature of the data is displayed in Figure 4.b. The figure contains only the trials without a dominating local winner, i.e. the trials in which both L and R win in exactly three frames each. We group the trials based on the similarity measure ΔS, and we control as usual for ΔV. Differently from the overall difference ΔV, the average difference ΔS consider the differences in each frame without the sign. After controlling for ΔV, we observe that a trial with low ΔS is comprised of values in each frame that are, on average, closer (and therefore more similar), with respect to an analogous trial with high ΔS. Trials with similar values (low ΔS, blue in Figure 4.b) are associated with higher accuracy, visible also as a steeper curve. Our definition of similarity arises from the characteristics of the pair of vectors that represents the option. A definition of *revealed* similarity as ease of comparison can be found in Natenzon (2019). The author introduces a stochastic choice model where similarity captures the correlation between perceptual errors, or preference shocks.

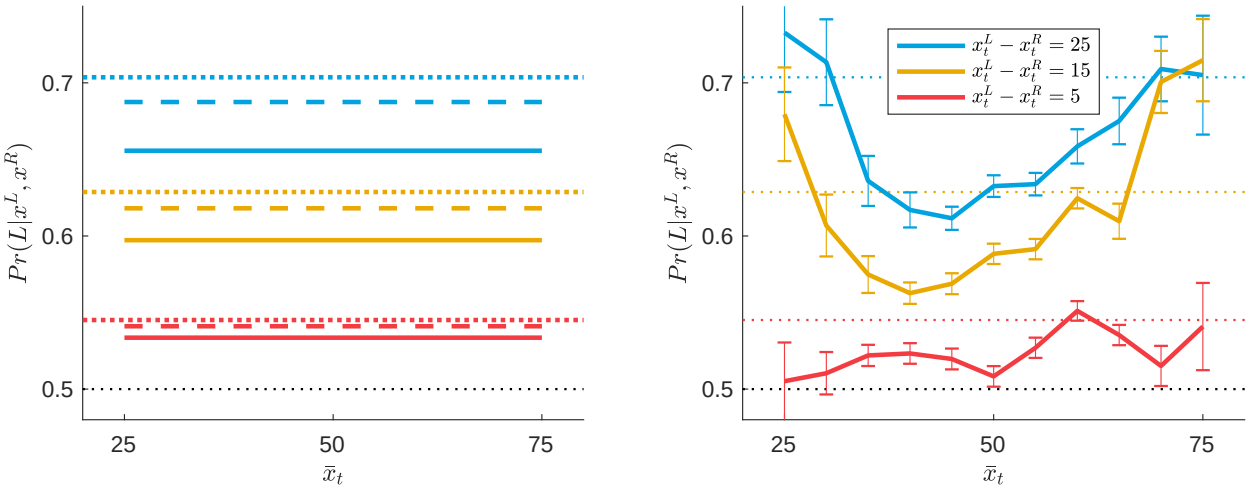


Figure 5: Decision weight $Pr(L|x_t^L, x_t^R)$ for different magnitudes $\bar{x}_t$ (x-axis) and differences Δx_t (color). (A) Simulations. Optimal observer (dotted lines), small variance (dashed lines), and high variance (solid lines). (B) Observational data (both treatments) grouped by Δx_t.

So far we focused on value differences within a frame and trial. We are interested in knowing whether *absolute* values have an additional effect on choices. In order to do so, we define the *decision weight* of a pair of values displayed in a frame. Each frame t of the trial contains two bars with values x_t^L and x_t^R. Values are independent across frames, therefore it is reasonable to expect that, ceteris paribus, a frame with $x_t^L > x_t^R$ increases the probability of choosing L over R, and such increase is higher when the within-frame difference is larger. The decision weight $Pr(L|x_t^L, x_t^R)$ associated with the pair $x_t^L > x_t^R$ is the probability that the final choice corresponds to the highest value in the current frame. We assume symmetricity between L and R, and this assumption seems reasonable given the analysis presented in Appendix B, where we run a robustness test in which we relax the symmetricity assumption.

A DM that perfectly observes and integrates the values would display decision weights that are increasing in the frame difference Δx_t, but do not depend on their absolute magnitude, nor on their average value $\bar{x}_t \equiv \frac{x_t^L + x_t^R}{2}$. The value difference is the only relevant piece of information, as shown by the simulations plotted in Figure 5 as dotted lines. Figure 5.a display two noisy versions of the DM above, that include a stochastic component with two different levels of variance. Details of the simulations are postponed to the next section, where we introduce different models of noisy perception. The absolute values are not relevant even in the case of noisy decision makers with constant variance in the perception of values. Only the difference determines the probability weights, that are monotonically increasing in Δx_t (compare lines across colors) and decreasing in the amount of noise we add (compare lines with the same color).

Observational data clearly do not display this property, and Figure 5.b shows an overall U-shaped pattern. This figure indicates that, on average, a difference of 20 points is weighted more in the choice when values are low (for example 15 and 35 percentage points) or high (e.g. 65-85), with respect to a case in which both values are closer to the average (e.g. 40-60). This puzzling pattern becomes clearer if we separate the observational data of the two treatments, displayed in Figure 6.

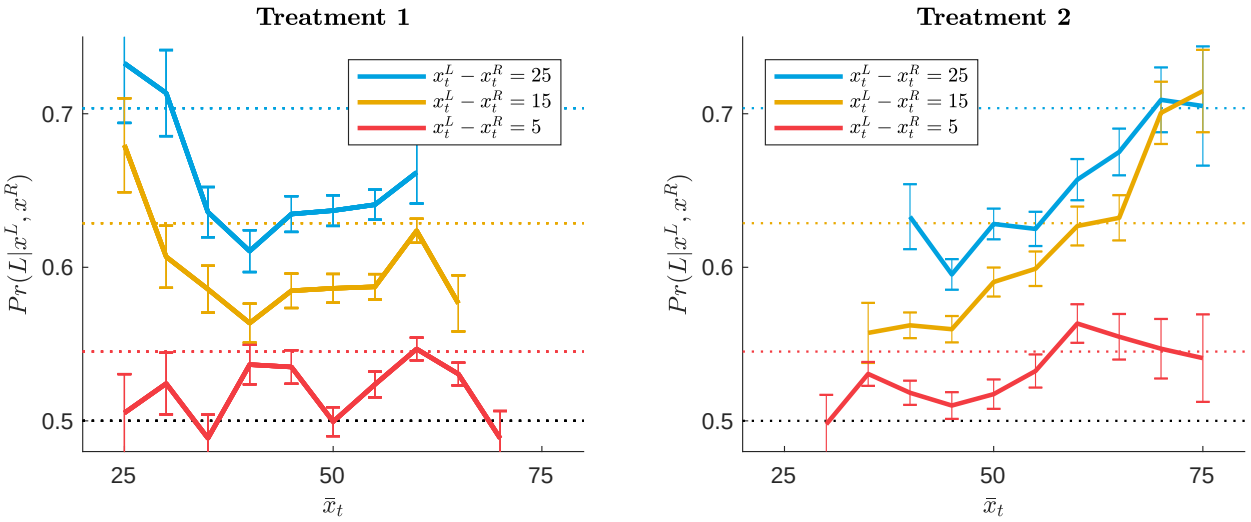


Figure 6: Observational data (separate treatments). Decision weight $Pr(L|x^L, x^R)$ for different magnitudes $\bar{x}_t$ and differences Δx_t.

The treatments differ in the data generating process of the stimuli, and in particular in which values are more common to observe. Treatment 1 (Figure 6.a) uses an upward triangular distribution of values, with low values being rare. Vice versa, treatment 2 (Figure 6.b) adopts a downward triangular distribution, with common low values and rare high values. By separating the treatments we observe the two components of the U-shaped pattern illustrated in Figure 5.b. Low [high] values have a high decision weight in treatment 1 [treatment 2], in which they appear less frequently. This pattern is consistent with the predictions of a *salience* model as the one introduced by Bordalo, Gennaioli, and Shleifer (2012). In their model, an agent observes the attributes of various products and, when she integrates the attributes, she overweights the most salient attribute for each product, i.e. the dimension that is more unusual given the observed distribution of data. Figure 6 suggests that the absolute value of the difference between the two values in each frame $|\Delta x_t|$ is overweight when the values displayed are unusual.

5 Model Fitting and Comparison

In this section we introduce various models that allow stochastic choice through noisy perception of the values. The first model that we consider is equivalent to a classic Random Utility Model with constant noise in the encoding of each value. We proceed by relaxing the assumptions of unbiased encoding and constant noise. After, we introduce direct comparison between pairs of values, instead of independent evaluation. Finally, we introduce *salience* in the model as focusing towards surprisingly high or low values. We compare the fit of the different models and we discuss which assumptions need to be relaxed in order to explain the choice patterns described in Section 4. Appendix B contains further restrictions on the aforementioned models.

5.1 Benchmark RU model: unbiased perception with constant noise

We consider a noisy perception model equivalent to the classic Random Utility Model with constant noise. Given a *true* (precise) displayed value x_t^i the DM perceives $\hat{x}_t^i = x_t^i + \epsilon_t^i$, a noisy representation of x_t^i, that includes also an additive measurement error $\epsilon \sim N(0, s^2)$. In our experiment the frames are shown sequentially, and we introduce a constant discount factor $0 < \delta \leq 1$ to capture leaking memory during the trial. The adoption of $\delta < 1$ corresponds to the introduction of a recency effect, with higher weight of late trials with respect to early ones. Appendix A contains evidence in favor of the adoption of a decaying exponential discount factor. The perceived value of option i is $\hat{X}^i$.

$$\hat{x}_t^i | x_t^i \sim N(x_t^i, s^2) \qquad \hat{X}^i = \sum_{t=1}^{T} \delta^{T-t} \cdot \hat{x}_t^i \tag{2}$$

We assume that there is no error in the decoding phase; the decision maker always chooses the option with the highest *perceived* value. Since the measurement errors are independent and normally distributed, we can calculate the choice probability given a pair of options:

$$Pr(\text{Choose } L | X^L, X^R, \delta, s^2) = Pr(\hat{X}^L > \hat{X}^R) = \Phi\left(\frac{\sum_{t=1}^{6} \delta^{6-t}(x_t^L - x_t^R)}{s \cdot \sqrt{2 \cdot \sum_{t=1}^{6} \delta^{6-t}}} \right) \tag{3}$$

where $\Phi(z)$ is the cumulative distribution function of a standard normal variable z. If $\delta = 1$ (no recency effect) and $s = 0$ (no perceptual noise) the DM is deterministic and chooses L if $^L > X^R$ (with ties broken randomly). As the perceptual noise increase, choice become more stochastic, up to $Pr(\text{Choose}L|\cdot) \to \frac{1}{2}$ regardless of the displayed values when $s \to \infty$.

5.2 Generalized noisy perception model (independent values)

The benchmark RU model assumes that the perception of each value is unbiased, i.e. $\mathbb{E}[\hat{x}|x] = x$, and the variance of the measurement error is constant and independent of the associated value $s^2(x) = s^2$. We generalize the model by relaxing both assumptions: the presented number is transformed by the function $m(x)$ and the noise can vary depending on the value $s^2(x)$.

$$\hat{x}_t^i | x_t^i \sim N(m(x_t^i), s^2(x_t^i)) \qquad \hat{X}^i = \sum_{t=1}^{T} \delta^{T-t} \cdot \hat{x}_t^i \tag{4}$$

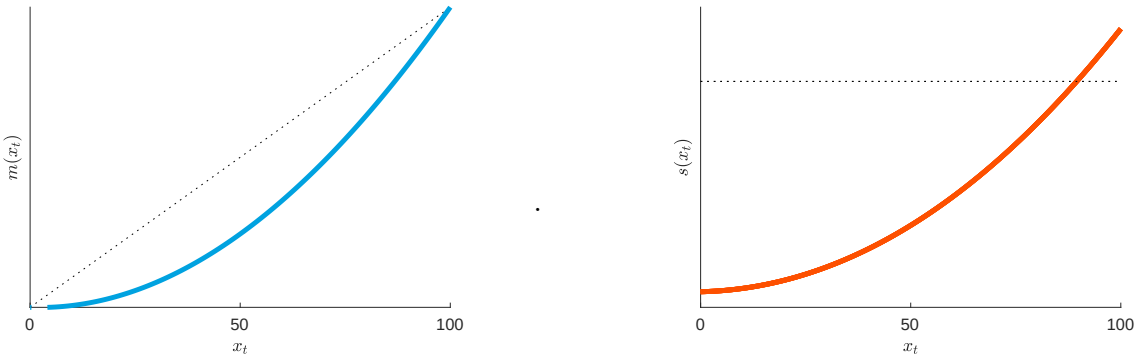
We maintain a constant discounting factor and deterministic choice conditional on perceived value, and we calculate the generalized choice probability:

$$Pr(\text{Choose } L | X^L, X^R, \delta, m(\cdot), s^2(\cdot)) = Pr(\hat{X}^L > \hat{X}^R) = \Phi\left(\frac{\sum_{t=1}^{6} \delta^{6-t}(m(x_t^L) - m(x_t^R))}{\sqrt{\sum_{t=1}^{6} \delta^{6-t}(s^2(x_t^L) + s^2(x_t^R))}} \right) \tag{5}$$

We are interested in providing a quantitative measure of how the data are more consistent with the flexible model (5) than with the more restrictive predictions of (3). We do this by finding the parameter estimates for each model that maximize the likelihood of the observed dataset. We consider the case of $m(x)$ and $s(x)$ being both polynomia of degree 3 with the restrictions over the standardized range $x \in [0, 1]$: $m(0) = 0, m(1) = 1$ and $s^2(x) > 0$.

The columns labeled LL in Table 1 reports the maximized value of the log-likelihood of the data (both separately for each treatment, and by merging the two sub-datasets). In the comparison between these models we expect the likelihood to be higher for the second, more flexible model, as the former is nested as a special case of the latter with $m(x) = x$ and $s^2(x) = s^2$. A more useful comparison is given by the Bayes information criterion (BIC) statistic, also reported in the table for each model, which penalizes the use of additional free parameters. This is defined as $BIC = k \cdot log(N_{obs}) - 2 \cdot LL$, where k is the number of free parameters for a given model, and N_{obs} is the number of observations in the dataset. The data provide more evidence in favor of the model with the lower BIC statistic. We can notice that the second model has a higher log-likelihood, but also higher BIC statistic, suggesting that the additional degrees of freedom provide little improvement in the fitting of data. Appendix B contains further restrictions on the second model, and show that the varying noise in particular penalizes the flexible model.
We display the difference between the two models visually in Figure 7. The plots show the fitted curves for $m(x)$ and $s(x)$ for the whole dataset, and the dotted lines indicate the fitted parameters for the first model. In the second model both the curves are monotonic and convex,

indicating disproportionate weight in the decision process of relatively high values, as well as higher noise in the encoding of high values.



Transformation $m(x)$ Varying noise $s(x)$

Figure 7: Fitted curves $m(x)$ and $s(x)$ for model 1 (dotted lines) and model 2 (solid lines).

5.3 Generalized noisy perception model (value differences)

We compare the second model with an analogous one in which value differences $\Delta x_t = x_t^L - x_t^R$ are encoded in each frame, instead of two independent values x_t^L and x_t^R.

$$\Delta\hat{x}_t|\Delta x_t \sim N(m(\Delta x_t^i), s^2(\Delta x_t^i)) \qquad \hat{X}^L - \hat{X}^R = \Delta\hat{X} = \sum_{t=1}^{T} \delta^{T-t} \cdot \Delta\hat{x}_t \tag{6}$$

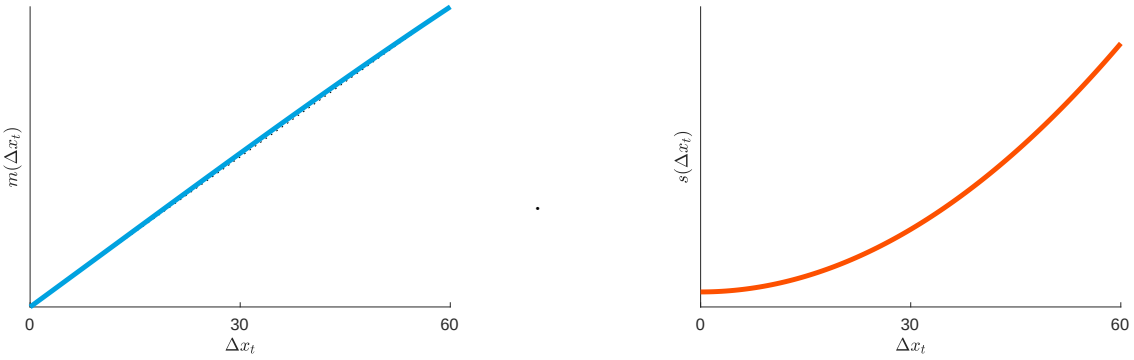
The functions $m(\Delta x_t)$ and $s(\Delta x)$ are polynomia of degree 3 with the restrictions over the standardized range $\Delta x \in [-1, 1]$: $m(0) = 0, m(-1) = m(+1) = 1$ and $s^2(x) > 0$. We add the symmetricity restriction $f(\delta x) = -f(-\delta x)$ for both the polynomia, and we proceed as before. The symmetricity restriction is discussed in Appendix B, where we provide evidence in favor of its plausibility. Now the choice probability explicitly depend on the *pairs* of values displayed in the same frame:

$$Pr(\text{Choose } L|X^L, X^R, \delta, m(\cdot), s^2(\cdot)) = Pr(\Delta\hat{X} > 0) = \Phi\left(\frac{\sum_{t=1}^{6} \delta^{6-t} m(x_t^L - x_t^R)}{\sqrt{\sum_{t=1}^{6} \delta^{6-t} s^2(x_t^L - x_t^R)}}\right) \tag{7}$$

Despite having the same number of degrees of freedom as the second model, Table 1 shows that this model achieves both an higher log-likelihood score and a lower BIC statistic with respect to both the previous models. We can interpret this result in favor of our hypothesis that values are jointly encoded through differences, instead of separately.

The noisy encoding of values differences seems even more promising if we observe the fitted

curves for $m(\Delta x)$ and $s^2(\Delta x)$ displayed in Figure 8. Remember that $m(x)$ is a polynomial of degree 3, but at first sight it is not possible to distinguish it from a linear curve. Imposing the restriction $m(\Delta x) = \Delta x$ further reduces the BIC statistic, as discussed in Appendix B. The variance is a monotonic and convex curve, indicating higher levels of noise associated with larger differences.



Transformation $m(\Delta x)$ Varying noise $s(\Delta x)$

Figure 8: Fitted curves $m(\Delta x)$ and $s(\Delta x)$ for model 3.

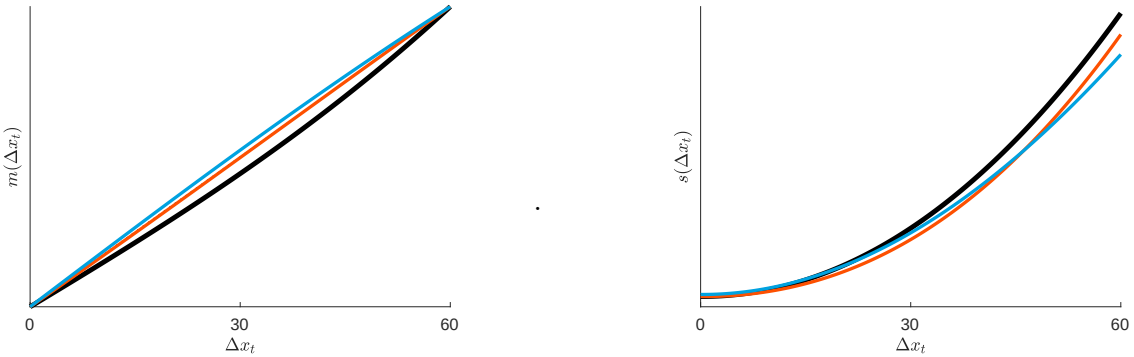
5.4 Noisy perception of differences and salience

A model of noisy perception of value differences as the one discussed above is able to capture the choice patterns displayed in Figure 4 (associated with dominating local winner and similarity between options), but not those in Figure 6, where we showed that the decision weights depend on the value magnitudes, additionally to their differences. In particular, unusually low or high values (depending on the treatment) are overweighted in the decision process. This pattern is consistent with the definition of *salience* adopted by Bordalo, Gennaioli, and Shleifer (2012). We extend the model introduced in the previous section to accommodate for different weights based on the average value currently presented in the frame.

$$\Delta\hat{x}_t|\Delta x_t \sim N(m(\Delta x_t^i)\cdot\bar{x}_t^{\mu}, s^2(\Delta x_t^i)\cdot\bar{x}_t^{\sigma}) \qquad \hat{X}^L - \hat{X}^R = \Delta\hat{X} = \sum_{t=1}^{T}\delta^{T-t}\cdot\Delta\hat{x}_t \tag{8}$$

where μ and σ capture the effect of the absolute values on decision weight and perception noise. The previous model is nested in the current one under the restriction $\mu = \sigma = 0$, whereas $\mu > 0$ [$\sigma > 0$] indicates larger weight [noise] associated with large magnitudes. We calibrate the model separately for each treatment, as well as for the whole dataset, as shown in Table 1. Additionally to the log-likelihood increase due to the additional parameters, we see a further decrease in the BIC statistics, confirming the relevance of the salience effect in the experimental dataset.

Figure 9 contains the fitted functions $m(\Delta x)$ and $s^2(\Delta x)$ for each treatments (red and blue lines), as well as for the entire dataset (black lines). The curves are not significantly different, suggesting that the fit improvement from analyzing the treatment separately is associated with the focusing parameters. In fact, we observe significant differences across treatments: calibrated values are $\mu = 0.99$ and 2.18 and $\sigma = 0.79$ and 1.95 for treatments 1 and 2 respectively.



Transformation $m(\Delta x)$ Varying noise $s(\Delta x)$

Figure 9: Fitted curves $m(\Delta x)$ and $s(\Delta x)$ for model 4: Treatment 1 (red lines), Treatment 2 (blue lines), All data (black lines).

Model	Merge T1 and T2		Separate T1 and T2	
	LL	BIC	LL	BIC
Noisy perception $N\left(x, s^2\right)$	-7,976	15,972	-7,965	15,969
Transformation of x $N\left(m(x), s(x)^2\right)$	-7.931	15,929	-7.882	15,900
Transformation of Δx $N\left(m(\Delta x), s(\Delta x)^2\right)$	-7,867	15,801	-7,846	15,826
Focusing based on $\bar{x}$ $N\left(m(\Delta x) \cdot \bar{x}^{\mu-1}, (s(\Delta x) \cdot \bar{x}^{\sigma-1})^2\right)$	-7,719	15,525	-7,675	15,524

Table 1: Summary table reporting the BIC for the main fitted models

Table 1 summarizes the results described in this section. We calibrate each of the four main models using the whole dataset, as well as separate treatments. For each of them we show log

likelihood (LL) and Bayesian Information Criterion (BIC) statistics. More results are shown in Appendix B.

6 Discussion and Concluding Remarks

We showed that, in a simple perceptual task, binary choice of multidimensional objects depends both on their true value difference and their similarity. We considered a class of models that gradually relaxes the assumptions underlying Random Utility Models with constant noise and allow value differences to be evaluated, instead of single values in isolation. Dimension-by-dimension comparison is able to explain the rich pattern of choices that we observe in the dataset (Figure 4), and highlight the role of ease of comparison between similar options. We fit the data using a model that allows perceptual noise to depend on the value difference on each dimension (Equations 6-7), and the fitted model shows unbiased perception of each dimension, and increasing and concave levels of noise (Figure 8). Small differences are encoded with higher accuracy than large ones. This result is consistent with the efficient coding hypothesis (Barlow 1961): higher accuracy (lower noise) is observed in correspondence of small differences, that appear more often in the dataset (Figure 2). Finally, we showed evidence of salience effect, defined as excessive weight in the decision process of unusual values. The adoption of two separate treatments with the same distribution of value differences, but different underlying absolute value distributions, generated different decision weights (Figure 6), a result that is at odd with the assumption that only relative values, and not absolute ones, enter in the decision process.

In this section we discuss how the model we introduced can be able to accommodate for violation of weak stochastic transitivity, and we suggest two possible applications. The applications serve a twofold purpose. On the one hand, offering decision making problem that can be nested in traditional strategic settings like consumer choice and voting behavior. On the other hand, emphasize the flexibility of the models we introduced, with major emphasis on the *bias* in the perception of differences (Application 1) or on the *noise* (Application 2). We conclude by highlighting the limitations and challenges of our approach, and emphasizing how further steps are required to extend the model to discrete choice models with more than two options.

6.1 Violations of Weak Stochastic Transitivity

Normative theories of decision making typically assume preferences to be rational, i.e. satisfy completeness and transitivity. When preferences map into choices in a deterministic way, transitivity means that given a triplet of options A, B, and C, if $A \succ B$ and $B \succ C$, it implies

that $A \succ C$. Empirical and experimental datasets consider stochastic choices, where different occurrences of the same choice problem may lead to different choices. In these cases the transitivity axioms must be relaxed, and it can be replaced with the more permissive weak stochastic transitivity (WST): if A is chosen more often over B, and B more often over C, it must be that A is chosen more often over C. When only binary choices occur, we can equivalently write that, under WST, if $Pr(A|\{A,B\}) > 0.5$ and $Pr(B|\{B,C\}) > 0.5$, it follows that $Pr(A|\{A,C\}) > 0.5$. Violations of WST occur when we observe a loop in the choices, for example when $Pr(A|\{A,B\}) > 0.5$, $Pr(B|\{B,C\}) > 0.5$, and $Pr(C|\{A,C\}) > 0.5$. Violations of WST are incompatible with RUMs with constant noise, as they predict that A is chosen more often than B if and only if the underlying value of the first option is higher than the latter, and values are real numbers that comply with transitivity.

The experimental literature offers extensive evidence of violation of WST (Tversky 1969, Regenwetter, Dana, and Davis-Stober 2011, Tsetsos et al. 2016) both at the individual and aggregate level, and we are interested in understanding whether DMs depart from transitivity following a consistent pattern. Our experimental design does not offer a systematic test of WST, as the values shown to participants are generated independently in each trial, but we see in Figure 4.a that choice probabilities depend on the number and direction of local wins, after controlling for the overall value difference. A model of noisy perception of value differences is able to predict violations of WST when three equally valued alternatives are compared pairwise, if they differ circularly in their number of winning dimensions and the DM adopts a biased transformation function $m(\Delta x) \neq \Delta x$.

To illustrate how violations of WST in binary choices can arise from noisy evaluation of differences, consider three equally valued alternatives displayed in Table 2. A, B, and C differ along three equally important dimensions, that are observed once at the time. In the binary choice between two options, for example A and B, we first define their difference in each dimension $\Delta x_t = x_t^A - x_t^B$. After, we apply a transformation $m(\Delta x_t)$; as an illustrative example we will consider the concave function $m(\Delta x_t) = sgn(\Delta x_t)\sqrt{|\Delta x_t|}$. Finally, the transformed differences are summed, and A is chosen if the sum is positive (suppose ties are broken randomly).
Since $m(\Delta x_t)$ is concave, the option with two winning dimensions over three will be chosen over an alternative that wins by a larger margin on a single attribute. When the same three values $(0, 0.5, \text{and } 1)$ are permuted circularly, the model predicts a violation of WST: A is chosen more often over B, B over C, and C over A.

	Options			A vs. B		B vs. C		C vs. A	
	A	B	C	Δx_t	$m(\Delta x_t)$	Δx_t	$m(\Delta x_t)$	Δx_t	$m(\Delta x_t)$
x_1^i	0	1	0.5	-1	-1	+0.5	+0.71	+0.5	+0.71
x_2^i	0.5	0	1	+0.5	+0.71	-1	-1	+0.5	+0.71
x_3^i	1	0.5	0	+0.5	+0.71	+0.5	+0.71	-1	-1
$\Sigma_t x_t^i$	1.5	1.5	1.5	0	+0.42	0	+0.42	0	+0.42
Choice				$A \succ B$		$B \succ C$		$C \succ A$	

Table 2: Qualitative predictions for all pairwise choices among three equally valued alternatives. We observe violation of weak stochastic transitivity as $A \succ B \succ C \succ A$.

A RUM with noisy encoding of each option in isolation would not be compatible with violations of WST: even if the encoding of each dimension is biased, for example with a concave transformation $m(x_t^i) = \sqrt{x_t^i}$, the *average* perceived value of each option will be independent from the other alternatives in the choice set, and after the introduction of perceptual noise it still implies that $Pr(\text{Choose } A|\{A, B\}) > 0.5 \Leftrightarrow v(A) > v(B) \Leftrightarrow \sum_t x_t^A > \sum_t x_t^B$.

Alternative explanations for the tendency to choose the frequent local winner option also require joint evaluation of the alternatives in the choice set. Normalization theories, such as divisive normalization (Louie, Khaw, and Glimcher, 2013) and range normalization (Soltani, De Martino, and Camerer, 2012), assume that dimensions are considered one at a time, like in our model, and original values are weighted with respect to the relative variance, or the overall range, of the dimension under consideration. A model of selective integration (Tsetsos et al. 2016) would also predict WST violations when the DM adopts an evaluation policy that discards information about alternatives with lower value in each dimension. The selective integration model can explain higher preference for options with high variance across own dimensions (provariance effect), but it is not able to predict the effect of similarity between alternatives on the ease of comparison, that was observed in the current experiment (Figure 4.b) and is consistent with our model when the noise variance is increasing in the observed difference.

6.2 Application 1: Buyer-seller product quality obfuscation

The first application we discuss is a simple buyer-seller scenario in which the buyer compares the products available one characteristic at the time, and chooses one of them based on the perceived value. The naive buyer is facing a binary choice between two products labeled by $i \in L, R$, with prices $p_i \geq 0$, $T \geq 1$ characteristics $X^i = \{x_t^i\}_t$, $x_t^i \in [0, 1]$, and *true* net value $v(X^i, p_i) = \sum_{t=1}^T x_t^i - p_i$. The buyer chooses the product with the highest *perceived* net value,

that does not necessarily coincide with the product with the highest true value. We define the perceived value difference $v(\Delta\hat{X})$ as the outcome of a noisy encoding of differences observed for each characteristics, whereas prices are encoded without noise. We follow closely the model described in Section 5.3, and we define

$$\Delta x_t = x_t^L - x_t^R \qquad \Delta\hat{x}_t|\Delta x_t \sim N(m(\Delta x_t^i), s^2(\Delta x_t^i)) \qquad v(\Delta\hat{X}) = \sum_{t=1}^{T} \Delta\hat{x}_t - (p_L - p_R)$$

As an illustrative example we will use $m(\Delta x) = sgn(\Delta x)(\Delta x)^2$ and $s^2(\Delta x) = (\Delta x)^2$.
It follows that

$$Pr(\text{Choose } L|X^L, X^R, p_L, p_R) = Pr(v(\Delta\hat{X}) > 0) = \Phi\left(\frac{\sum_{t=1}^{T} sgn(x_t^L - x_t^R)(x_t^L - x_t^R)^2 - p_L + p_R}{\sqrt{\sum_{t=1}^{T}(x_t^L - x_t^R)^2}}\right) \tag{9}$$

By fixing a seller L who offers a product with characteristics X^L and price p_L, the competitor R can design an alternative product that is overall worse for the buyer (i.e. $\sum_{t=1}^{T} x_t^R < \sum_{t=1}^{T} x_t^L$) and is sold for a higher price $p_R > p_L$, and still achieve a market share greater than half the market. Consider for example $X^L = \{0.5, 0.5, 0.5\}$ (tot 1.5), $X^R = \{0.2, 0.2, 1\}$ (tot 1.4), $p_L = 0$, $p_R = 0.01$. By following Equation 9 we obtain

$$Pr(\text{Choose } R) = \Phi\left(\frac{0.25 - 0.09 - 0.09 - 0.01}{\sqrt{0.09 + 0.09 + 0.25}}\right) \approx 0.56 > 0.5$$

In this example $m(\Delta x)$ is convex and the buyer is biased in favor of products that are much better in few characteristic. Choice reversal is possible both when $m(\Delta x)$ is convex and concave. In the latter case, the competitor would design own product's characteristics in order to appear slightly better in most of the dimensions, and significantly worse in one dimension only. The introduction of a bias in the transformation function $m(\Delta x)$ is necessary in order to obtain choice reversal (with respect to true preference). In the next application we provide an example in which, without any systematic bias, the difficulty in the comparison between extremely different options make the decision maker less accurate, and therefore more likely to make mistakes, although not enough to observe choice reversal.

6.3 Application 2: Voter-party policy platform design

The second application we introduce is a voting scenario in which a single voter compares two party platforms, one policy at the time, and votes the party whose *perceived value* is higher. Each party $i \in L, R$ offers a policy platform over $T \geq 1$ issues $X^i = \{x_t^i\}_t$, $x_t^i \in [0, 1]$. The voter

evaluates the party with respect to an ideal platform $Y = \{y_t\}_t$. Voter's true preferences are linear with respect the distance between the ideal and the implemented platform, and the value associated to each platform can be expressed as $v(X^i, Y) = \sum_{t=1}^{T} -|x_t^i - y|$. She votes the party whose platform has the highest *perceived* value, that does not necessarily coincide with the one with the highest true value. We define the perceived value difference $v(\Delta_{(y)}\hat{X})$ as the outcome of a noisy encoding of differences observed for each topic. We adapt the model described in Section 5.3 appropriately, and we define

$$\Delta_{(y)}x_t = |x_t^R - y_t| - |x_t^L - y_t| \qquad \Delta_{(y)}\hat{x}_t|\Delta_{(y)}x_t \sim N(m(\Delta_{(y)}x_t^i), s^2(\Delta_{(y)}x_t^i))$$

$$v(\Delta_{(y)}\hat{X}) = \sum_{t=1}^{T} \Delta_{(y)}\hat{x}_t$$

As an illustrative example we will use $m(\Delta_{(y)}x) = \Delta_{(y)}x$ and $s^2(\Delta_{(y)}x) = (\Delta_{(y)}x)^2$.
It follows that

$$Pr(\text{Vote } L|X^L, X^R, Y) = Pr(v(\Delta_{(y)}\hat{X}) > 0) = \Phi\left(\frac{\sum_{t=1}^{T} |x_t^R - y_t| - |x_t^L - y_t|}{\sqrt{\sum_{t=1}^{T}(x_t^L - x_t^R)^2}}\right) \tag{10}$$

By fixing an incumbent party L with a platform X^L, the challenger R can design own platform in order to facilitate or impede comparison. We abstract from constraint and motivation for the challenger party in the design on the platform, and we focus on a series of examples in which the voter equally dislikes the challenger, but votes for them with different probabilities based on the ease of comparison between platforms.

Suppose that the ideal platform for the voter is moderate in each issue, with $Y = \{0.5, 0.5, 0.5\}$, and the incumbent's platform is $X^L = \{0.25, 0.25, 0.25\}$.
It follows that $v(X^L, Y) =_{t=1}^{3} -|x_t^L - y| = -3 \cdot |0.25 - 0.5| = -0.75$. We introduce three different cases in which $v(X^R, Y) = -0.9 < v(X^L, Y)$, but the voting probability is different.
First, consider the case in which the parties are identical in two out of three topics, and differ only in the third one: $X^{R1} = \{0.1, 0.25, 0.25\}$. Comparison is very easy, and in fact (under the functional forms assumed above) $Pr(\text{Vote } L|X^L, X^{R1}, Y) = \Phi\big(\frac{0.15}{\sqrt{0+0+0.15^2}}\big) \approx 0.84$.
Suppose now that the challenger wants the comparison to be less obvious, by taking a moderate position in one topic, and a more extreme one in the other two topics: $X^{R2} = \{0.4, 0.1, 0.1\}$. Comparison is less intuitive, and in fact $Pr(\text{Vote } L|X^L, X^{R2}, Y) = \Phi\big(\frac{-0.15+0.15+0.15}{\sqrt{0.15^2+0.15^2+0.15^2}}\big) \approx 0.72$.
Finally, suppose the challenger wants the comparison to be extremely difficult, and she maximizes the distance from the incumbent by selecting more extreme positions on the other side of

the available spectrum, for instance $X^{R3} = \{0.8, 0.8, 0.8\}$. The two platforms are now dissimilar and difficult to compare, and the noisy voter will choose the preferred option with a probability barely above chance: $Pr(\text{Vote } L|X^L, X^{R3}, Y) = \Phi\big(\frac{3 \cdot 0.05}{\sqrt{3 \cdot 0.55^2}}\big) \approx 0.56$.

6.4 Limitations

The experimental design and the models we considered to fit the data rely on a few assumptions that deserve to be discussed further. First, dimension-by-dimension comparison is facilitated in the experiment, as the participant did not have the opportunity to evaluate the two options in separate phases. Other experimental paradigms that allow participants to freely explore the dimensions in the domain of choice under risk (Lohse and Johnson 1996), strategic interactions (Polonio, Di Guida, and Coricelli 2015), and product choice (Gidlof et al. 2013) show that humans adopt heterogeneous explorations (and comparison) strategies, and these are correlated with the choice behavior. Despite context effects are observable even when comparison is not enforced, the magnitude of the effect we found in the data may be inflated by this specific feature of the perceptual task. On a similar direction, values in the experiment were displayed as bars, whose length is proportional to the underlying value. It is not obvious whether this feature should enhance or reduce the magnitude of the effect with respect to the adoption of numerical values. The adoption of visual stimuli is an intuitive way to represent dimensions and add perceptual noise in the evaluation of each value in isolation. Yet our task rely on comparisons, and the difference between bar lengths (that is perceived visually) may be encoded differently with respect to the difference between numerical values (that require a cognitive process to perform the subtraction). Comparing our results with an analogous design that uses Arabic numerals will be helpful to shed light on this point.

The models of noisy perception of differences we discussed are inherently grounded on the assumption that only two options are compared at the time. A further generalization should redefine how comparison occurs when $N > 2$ options are available, for example in a trinary choice between two alternatives and an outside option (as in the case of purchase choice between products A and B, or no purchase). We believe that such a generalization would be valuable to apply a model of noisy perception based on comparison to a vast range of discrete choice problems where RU with constant noise represents the benchmark model. Going beyond a mere theory-testing motivation, econometric methods can be generalized with lighter restrictions on the structure of the noise. The adoption of a perceptual noise, instead of a preference shock, also has deep empirical implications for efficiency and welfare. Matejka and McKay (2015) provided a microfoundation to the discrete-choice RUM based on costly and incomplete information acquisition, using a Rational Inattention (RI) framework. Although RI and RUM frameworks may

yield identical choice probabilities, welfare-calculation are different because of the information costs and the nature of the stochasticity in the evaluation of each option. Joo (2019) showed that in a model with imperfect information the introduction of a bad alternative to the choice set may reduce overall welfare, whereas this is not possible in traditional RUMs. Similarly, we showed in the two applications that a strategic agent that is able to manipulate the choice set can reduce the DM's expected value by taking advantage of the perceptual noise and the difficulty in comparing options that differ in multiple dimensions. Exploring further the interaction between perceptual noise in the evaluation process and endogenous manipulation of choice sets, and its implication on welfare, is an interesting direction which we hope to pursue in future research.

References

Barlow, H.B., 1961. Possible principles underlying the transformation of sensory messages. Sensory communication, 1, pp.217-234.

Block, H.D. and Marschak, J., 1960. Random orderings and stochastic theories of response, in Contributions to Probability and Statistics(I. Olkin, S. Ghurye, W. Hoeffding, W. Madow, and H. Mann, Eds.).

Bordalo, P., Gennaioli, N. and Shleifer, A., 2012. Salience theory of choice under risk. The Quarterly journal of economics, 127(3), pp.1243-1285.

Brown, J., Hossain, T. and Morgan, J., 2010. Shrouded attributes and information suppression: Evidence from the field. The Quarterly Journal of Economics, 125(2), pp.859-876.

Bushong, B., Rabin, M. and Schwartzstein, J., 2017. A model of relative thinking. Unpublished manuscript, Harvard University, Cambridge, MA.

Chetty, R., Looney, A. and Kroft, K., 2009. Salience and taxation: Theory and evidence. American economic review, 99(4), pp.1145-77.

Ellison, G., 2005. A model of add-on pricing. The Quarterly Journal of Economics, 120(2), pp.585-637.

Gabaix, X. and Laibson, D., 2006. Shrouded attributes, consumer myopia, and information suppression in competitive markets. The Quarterly Journal of Economics, 121(2), pp.505-540.

Gidlof, K., Wallin, A., Dewhurst, R. and Holmqvist, K., 2013. Using eye tracking to trace a cognitive process: Gaze behaviour during decision making in a natural environment.

Joo, J., 2019. Rational Inattention as an Empirical Framework-With an Application to the Welfare Effects of New Product Introduction and Endogenous Promotion. Available at SSRN 3182795.

Khaw, M.W., Li, Z. and Woodford, M., 2017. Risk aversion as a perceptual bias (No. w23294). National Bureau of Economic Research.

Kivetz, R., Netzer, O. and Srinivasan, V., 2004. Extending compromise effect models to complex buying situations and other context effects. Journal of Marketing Research, 41(3), pp.262-268.

Koszegi, B. and Szeidl, A., 2012. A model of focusing in economic choice. The Quarterly journal of economics, 128(1), pp.53-104.

Landry, P. and Webb, R., 2019. Pairwise normalization: A neuroeconomic theory of multi-attribute choice. Available at SSRN 2963863.

Li, V., Castanon, S.H., Solomon, J.A., Vandormael, H. and Summerfield, C., 2017. Robust averaging protects decisions from noise in neural computations. PLoS computational biology, 13(8), p.e1005723.

Lohse, G.L. and Johnson, E.J., 1996. A comparison of two process tracing methods for choice tasks. Organizational Behavior and Human Decision Processes, 68(1), pp.28-43.

Louie, K., Khaw, M.W. and Glimcher, P.W., 2013. Normalization is a general neural mechanism for context-dependent decision making. Proceedings of the National Academy of Sciences, 110(15), pp.6139-6144.

Luce, R.D., 1959. On the possible psychophysical laws. Psychological review, 66(2), p.81.

Matejka, F. and McKay, A., 2015. Rational inattention to discrete choices: A new foundation for the multinomial logit model. American Economic Review, 105(1), pp.272-98.

Mosteller, F. and Nogee, P., 1951. An experimental measurement of utility. Journal of Political Economy, 59(5), pp.371-404.

Natenzon, P., 2019. Random choice and learning. Journal of Political Economy, 127(1), pp.419-457.

Piccione, M. and Spiegler, R., 2012. Price competition under limited comparability. The quarterly journal of economics, 127(1), pp.97-135.

Polonio, L., Di Guida, S. and Coricelli, G., 2015. Strategic sophistication and attention in games: An eye-tracking study. Games and Economic Behavior, 94, pp.80-96.

Regenwetter, M., Dana, J. and Davis-Stober, C.P., 2011. Transitivity of preferences. Psychological review, 118(1), p.42.

Sinaiko, A.D. and Hirth, R.A., 2011. Consumers, health insurance and dominated choices. Journal of Health Economics, 30(2), pp.450-457.

Soltani, A., De Martino, B. and Camerer, C., 2012. A range-normalization model of context-dependent choice: a new model and evidence. PLoS computational biology, 8(7), p.e1002607.

Spiegler, R., 2006. Competition over agents with boundedly rational expectations. Theoretical Economics, 1(2), pp.207-231.

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

Thurstone, L.L., 1927. A law of comparative judgment. Psychological review, 34(4), p.273.

Tsetsos, K., Moran, R., Moreland, J., Chater, N., Usher, M. and Summerfield, C., 2016. Economic irrationality is optimal during noisy decision making. Proceedings of the National Academy of Sciences, 113(11), pp.3102-3107.

Tversky, A., 1969. Intransitivity of preferences. Psychological review, 76(1), p.31.

Tversky, A. and Simonson, I., 1993. Context-dependent preferences. Management science, 39(10), pp.1179-1189.

Weber, E.H., 1978. The Sense of Touch. New York: Academic Press [English translation of German original, 1834.]

Appendix A - Robustness Checks for Data Quality

Participants are randomly assigned to the treatments. Subjects registered with an odd number are assigned to Treatment 1 (upward triangular distribution), whereas subjects with an even number are assigned to Treatment 2 (downward triangular distribution). The two groups seem sufficiently similar in terms of overall accuracy during the first task of the experiment.

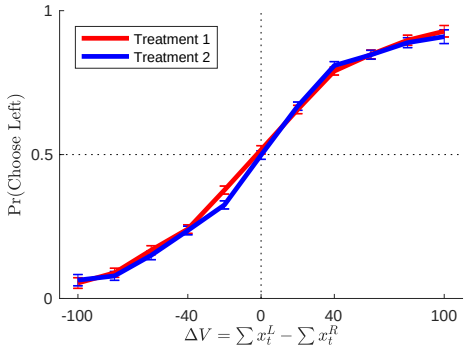


Figure 10: Choice probability in trials with different difficulty. Observations are grouped by treatment. Avg. accuracy 73.03% vs 74.36%.

Learning is a common concern in the experimental literature. Despite extensive instructions, practice trials, and information about the data generation process, we want to see if participants' accuracy is changing during the session. If we compare early trials (1-150) and late trials (151-300) we see a slight decrease in the average accuracy. This pattern is consistent with boredom, and it is less concerning than the opposite one. We can conclude that there is no strong *learning* due to additional information dismissed in the instructions phase.

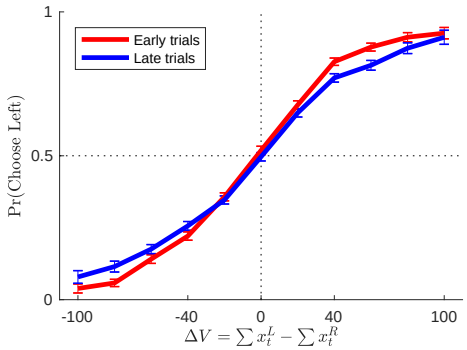


Figure 11: Choice probability in trials with different difficulty. Observations are divided into early (1-150) and late trials (151-300). Avg. accuracy diminishes from 74.71% to 72.65%.

Along the paper we adopt the parsimonious assumption that leaking memory is modeled with an exponential discounting factor, such as the values displayed at time t are discounted by $0 < \delta < 1$ with respect to the values displayed at time $t + 1$ (with $t = 6$ being the last frame displayed on the screen during the experiment). Figure 12 shows that this is represents a good approximation of the relative importance of values shown at different points in the trial. The blue curve indicates the discounting factor of each frame when calibration is free (6 values with unitary sum, for a total of 5 degrees of freedom). The red curve shows the calibration of the discount factor under a single parameter under the assumption of exponential discounting.

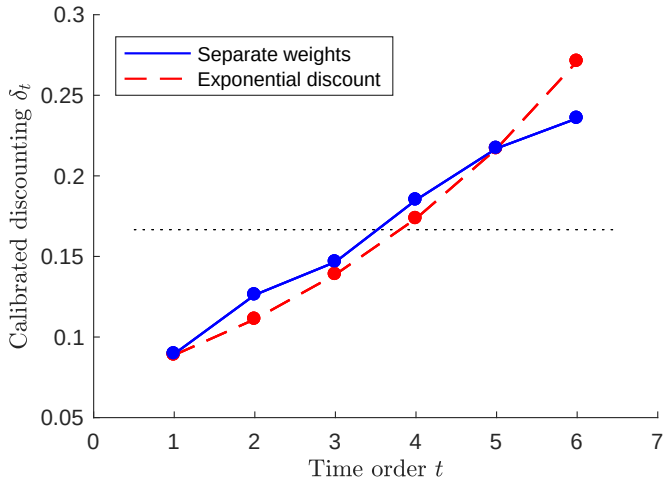


Figure 12: Calibrated discounting factor, here normalized in order to sum up to 1. Blue: Six separate weights are calibrated based on the order of appearance of values on the screen. Red: A single parameter is calibrated under the assumption of exponential discounting of past information. Dotted horizontal line: benchmark case without discounting, equal (normalized) weights $\delta_t = 1/6$.

Appendix B - Additional Fitted Models

In this section of the Appendix we consider an array of variations of the four main models introduced in Section 5. We include restrictions and generalizations in order to provide further evidence of the adopted assumptions, as well as which generalizations contribute improve the fit of the dataset. Table 3 summarizes log-likelihood and Bayesian Information Criterion for each of the models listed below.

Unless differently specified, we assume that value integration in Models 1 and 2 is

$$\hat{X}^L - \hat{X}^R = \Delta\hat{X} = \sum_{t=1}^{6} \delta^{6-t} \cdot (\hat{x}_t^L - \hat{x}_t^R)$$

and in Models 3 and 4 are

$$\hat{X}^L - \hat{X}^R = \Delta\hat{X} = \sum_{t=1}^{6} \delta^{6-t} \cdot \Delta\hat{x}_t$$

and choice is deterministic given the perceived value difference

$$Pr(\text{Choose } L|\Delta\hat{X} \geq 0) = 1 \quad \text{and} \quad Pr(\text{Choose } L|\Delta\hat{X} < 0) = 0$$

Model 1.0 - Equations 2-3 (Noisy perception)

Model 1.1 - Remove the assumption of exponential discounting, and allow the discount factor to vary freely across time (see also Figure 12 in Appendix A)

Model 2.0 - Equations 4-5 (Transformation of x)

Model 2.1 - Perceptual noise is constant: add the constraint $s(x) = s$

Model 2.2 - Unbiased perception: add the constraint $m(x) = x$

Model 2.3 - Relax the symmetricity assumption: define separately two cases $m_L(x_L)$ and $m_R(x_R)$, such that the DM can have a systematic bias in favor of one of the sides of the screen

Model 3.0 - Equations 6-7 (Transformation of Δx)

Model 3.1 - Perceptual noise is constant: add the constraint $s(\Delta x) = s$

Model 3.2 - Unbiased perception: add the constraint $m(\Delta x) = \Delta x$

Model 3.3 - Relax the symmetricity assumption: define separately two cases $m_L(\Delta x)$ when $\Delta x \geq 0$ and $m_R(\Delta x)$ when $\Delta x < 0$, such that the DM can have a systematic bias in favor of positive differences on one of the side of the screen

Model 4.0 - Equations 7-8 (Focusing based on $\bar{x}$)

Model 4.1 - Perceptual noise is constant: add the constraint $s(\Delta x) = s$

Model 4.2 - Unbiased perception: add the constraint $m(\Delta x) = \Delta x$

Model 4.3 - Relax the symmetricity assumption: define separately two cases $m_L(\Delta x)$ when $\Delta x \geq 0$ and $m_R(\Delta x)$ when $\Delta x < 0$, such that the DM can have a systematic bias in favor of positive differences on one of the side of the screen

Model (k = degrees of freedom)	Merge T1 and T2		Separate T1 and T2	
	LL	BIC	LL	BIC
Noisy perception				
Model 1.0 (k=2)	-7,976	15,972	-7,965	15,969
Model 1.1 (k=6)	-7,971	16,000	-7,962	16,040
Transformation of x				
Model 2.0 (k = 7)	-7.931	15,929	-7.882	15,900
Model 2.1 (k = 4)	-7.970	15,979	-7.921	15,919
Model 2.2 (k = 5)	-7.937	15,913	-7.890	15,876
Model 2.3 (k = 9)	-7.930	15,947	-7.879	15,931
Transformation of Δx				
Model 3.0 (k = 7)	-7,867	15,801	-7,846	15,826
Model 3.1 (k = 4)	-7,929	15,897	-7,928	15,933
Model 3.2 (k = 5)	-7,870	15,779	-7,851	15,798
Model 3.3 (k = 9)	-7.865	15,817	-7,842	15,857
Focusing based on $\bar{x}$				
Model 4.0 (k = 9)	-7,719	15,525	-7,675	15,523
Model 4.1 (k = 6)	-7,868	15,794	-7,843	15,802
Model 4.2 (k = 7)	-7,721	15,509	-7,678	15,491
Model 4.3 (k = 11)	-7,715	15,536	-7,673	15,558

Table 3: Summary table reporting the BIC for the main fitted models and their variations.

Appendix C - Experimental Instructions

The next pages contain the transcript of the instructions for the experimental session and the main task (please note that two ancillary tasks were performed in the second part of the session). The text was displayed on the screen in several pages, and each participant had the chance to read the instructions freely and go back to previous pages. The instructions were accompanied by examples and screenshots (described below), and included several practice trials (also described below). When instructions differ across treatments, the text is written in *italics* for treatment 1 and in [brackets] for treatment 2.

Welcome to CELSS: Columbia Experimental Laboratory for Social Sciences! Thank you for taking part in the experiment. Click on the arrows to proceed across the instruction pages. If something is unclear, raise your hand and ask for assistance.

During this game, you will play the role of a king who builds an army of knights to defend the castle from a wave of enemies. Orcs, trolls, and other monsters will attack you sequentially. Select the strongest knights to increase the chance of defending your castle!

The game consists of a perceptual task. In each round, you will get to pick between two knights, whose strength is represented by a sequence of bars with different lengths. You are asked to choose a knight, who will fight against a random enemy. Your victory in the round depends on a combination of skill and luck. Your objective is to win as many fights as possible. The final payment will depend on the number of victories (more details ahead). In order to achieve your goal, you should try to choose the knight with the highest average chance of winning against the different types of opponents.

This experiment will last for about one hour (800 rounds). Try to choose the best option in each round. On your side you have an army of potential knights. They differ in their ability to fight various enemies. For example, one could be strong against trolls but weak against goblins. The following instructions apply to the first 300 rounds, with minor changes later.

Consider the first example. You can choose between two knights, located on the left and right sides of the screen. Each knight is displayed in a column and has a set of six skills. Each row contains a skill and indicates the probability of winning against enemies of a specific type: orc, troll, dragon, etc. expressed in percentage points. Enemy types spawn with equal probability.

[Image: Screenshot of a simplified practice trial, with all the bars visible at once]

After the information disappears, you have 10 seconds to select a knight. Use the keyboard and press F to select the knight on the left or J to select the knight on the right. You will immediately receive a visual cue that indicates whether you won the round. A colored cross will appear at the center of the screen: a green cross indicates victory, and a red cross indicates defeat. Every time you win a round, you collect a victory point. Points are added through the whole experiment and determine your reward! Every few rounds you can check the total number of points. Click the left arrow icon to repeat the instructions, or the right arrow icon to proceed with the tutorial (no victory point will be added for now).

[Two simplified practice trials, with all the bars visible at once]

In the rest of the tutorial, and in the whole experiment, information will be shown sequentially on the screen: each skill is displayed for a fraction of second as a vertical bar (without numbers), and then the next skill will appear, for a total of six times per round. It is going to be more difficult than before! Let‘s try now a SLOW round. Click the arrow icon to proceed.
[Image: Screenshot of a experiment trial, with bars visible sequentially]

[A practice trial with slower speed - 1 second per frame]

During each round you have 10 seconds to select a knight. If you do not act on time, one knight will be selected randomly. Try to avoid this! The game will last about one hour, with blue “pause“ screens to take breaks every few minutes. Overall there are 800 quick rounds, grouped into blocks. Your final payment increases by 20 cents for every victory point after the first 300. If you get 375 victories you will receive $15, if you get 376 you will receive $15.20, and so on. In general, if you have X victory points you will receive (X-300)*20cents. You are guaranteed to receive at least $10 as your average success rate after a few rounds cannot be less than 45 %.

Your final payment will depend on your skill and luck. If you are choosing randomly, you have only a 50% probability of getting more than $20. If you are perfectly accurate in your choices, there is still a luck component but you have a 97% probability of winning more than $20, and if you are accurate 4 times out of 5 you will have a 87% probability of winning more than $20. Click the arrow to try another practice round.

[Another practice trial with slower speed - 1 second per frame]

Click the left arrow icon to repeat the practice rounds, or the right arrow icon to proceed with the tutorial.

You are now ready to learn more about the knights‘ strength. The pairs of knights are randomly generated by the computer, and each bar height (i.e. the strength, or winning probability) is independent from the others. First, the strength of a knight does not depend on the alternative knight on the screen. Second, the strength of a knight against a certain enemy does not depend on the strength against a different type of enemy. Finally, the strength levels lie in the range *between 10 and 70, with higher values more common than lower ones* [between 30 and 90, with lower values more common than higher ones].

Each knight‘s strength against each possible opponent is an independent random draw from the distribution of possible values shown below. Note that the frequency distribution is triangular, with *higher values more common than lower ones* [lower values more common than higher ones].
[Image: Frequency of values - upper or lower triangular distribution]

Regardless of your statistics training, you have now the opportunity to practice with the values that will appear during the rest of the experiment. After you click the arrow, you will see a fast sequence of bars randomly generated by the computer. This will help familiarize you with the values before collecting points. You can spend as much time as you want observing these bars. When you feel ready, click the arrow and proceed with the last few normal-speed practice rounds.

[Free learning phase - participants observe a sequence of values to familiarize with the task]

You will now face five normal-speed practice rounds. These five practice rounds have the normal speed that will be used for the whole game, but they are still not counted for the final score.

[Five practice trial with normal speed]

You have successfully completed the practice rounds. If you have any question about the game, you can ask the experimenter. When you feel ready to start the rewarded rounds, click the arrow icon.