University of Zurich

Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 330

Strength of Preference and Decision Making Under Risk

Carlos Alós-Ferrer and Michele Garagnani

July 2019

Strength of Preference and Decision Making Under Risk*

Carlos Alós-Ferrer†[1] and Michele Garagnani[1]

[1]Department of Economics, University of Zurich. Blümlisalpstrasse 10, CH-8006 Zurich, Switzerland

This Version: July 19th, 2019

Abstract

Overwhelming evidence from the cognitive sciences shows that, in simple discrimination tasks (determining what is louder, longer, brighter, or even which number is larger) humans make more mistakes and decide more slowly when the stimuli are closer along the relevant scale. We investigate to what extent these effects are relevant for economic decisions. Strikingly, we find that even when there is an objectively correct answer independently of attitudes toward risk, the same effects obtain as expected values become closer. Contrary to pure discrimination tasks, however, differences in payoff-independent numerical magnitudes play a minor role. When correct answers depend on subjective attitudes toward risk, differences in expected values fail to explain error rates. The gradual effects on error rates and response times subsist but are instead explained by cardinal differences in independently-estimated subjective utilities ("strength of preference"). This is in agreement with assumptions typically made (but seldom validated) in random utility models. We conclude that the gradual effects on choice found in cognitive discrimination paradigms are very much present in economic choices, but depend on purely economic variables. An implication is that even if correct economic choices can be seen as ordinal, actual economic choices carry a cardinal component.

JEL Classification: D9 · D01 · D81

Keywords: Strength of preference · Choice difficulty · Stochastic choice · Risk attitude

*The authors thank Ernst Fehr, Nick Netzer, Antonio Rangel, and Michael Woodford for helpful comments.

†Corresponding author: carlos.alos-ferrer@econ.uzh.ch

1 Introduction

Errors are everywhere. Economics has by now embraced the view that economic choices are subject to noise (e.g., McFadden, 2001). Research in stochastic choice has provided extensive evidence that human beings often make different choices even when repeatedly confronted with the same set of options[1] (e.g., Tversky, 1969; Camerer, 1989; Hey and Orme, 1994; Ballinger and Wilcox, 1997; Agranov and Ortoleva, 2017). There is, however, no universally-accepted view on the origins and determinants of noise or errors in economic decision making. How often do economic agents make mistakes, and what does the number of mistakes depend on? These questions are important both for positive and normative reasons. On the one hand, forecasting economic choices requires accurate models of decision errors, beyond the simple assertion that people indeed do make mistakes. On the other hand, predicting the effects of economic policies and evaluating their consequences is only possible if the consequences of human errors in response to them are understood. Indeed, large individual error rates are reflected in significant behavioral heterogeneity and can cause potentially large welfare losses at the aggregate level (e.g., Choi et al., 2014; Harrison and Ng, 2016; Alekseev et al., 2019).[2]

The key question is whether error rates are associated with directly or indirectly measurable *economic* variables. To understand the sources of mistakes (or stochastic choice) in economic decisions, however, it is useful to briefly step back and examine evidence from the cognitive sciences (chiefly cognitive psychology and neuroscience) on tasks which are significantly simpler than the ones proper of economics. In the domain of *psychophysics*, decades of research have concentrated on *perceptual discrimination tasks*, where two stimuli are presented and human participants are asked to estimate which one scores higher along an objective scale, for instance which of two sounds is louder, which of two lights is brighter, or which of two lines is longer. In such simple tasks, there is an objective, direct measure of choice difficulty: choices become gradually harder as the difference between the stimuli becomes smaller (along the objective scale). There are two firmly-established stylized facts in this literature. The first is that the percentage of correct choices is strictly decreasing with choice difficulty, that is, error rates are larger when stimuli are more similar (Laming, 1985; Klein, 2001; Wichmann and Hill, 2001). The second is that choices are slower as choice difficulty increases, that is, response times are larger when the stimuli are more similar (Dashiell, 1937; Moyer and Landauer, 1967). This second fact is commonly taken as evidence that the effect on error rates derives from basic (gradual) neural mechanisms in the human brain. That is,

[1] "Common experience suggests, and experiment confirms, that a person does not always make the same choice when faced with the same options, even when the circumstances of choice seem in all relevant aspects to be the same." (Davidson and Marschak, 1959).

[2] If a normative view is adopted where (except for knife-edge indifference cases) only one choice is considered correct (or consistent with underlying preferences), the statement that choice is stochastic is equivalent to the empirically-ubiquitous observation of positive error rates. It is in this sense that we speak of "errors" in this work. This is also in line with a positive-economics view, where one aims to understand the extent to which economic decision makers will deviate from choices deemed "rational."

decisions might derive from gradual, noisy processes of internal evidence accumulation, which are more error-prone and time-consuming if the quantities that need to be teased apart are closer (Shadlen and Kiani, 2013; Shadlen and Shohamy, 2016).

In this work, we ask the question of whether these gradual effects are relevant for economic choices and, if so, which economic variables do determine them. This question is obviously important for conceptual reasons, as the phenomena we discuss imply a *cardinal* effect of economic variables on choices, as opposed to the classical, purely-ordinal view of preferences. Providing an empirical demonstration of the postulated effects, however, is also important in view of the recent literature. Such effects would be an implication of any model assuming that choice frequencies reflect an efficient use of limited representational resources in the human mind, as for instance recent models of rational inattention (Caplin and Dean, 2015; Matějka and McKay, 2015) or optimal sparsity (Gabaix et al., 2006; Gabaix, 2014). Also, if certain differences among alternatives are more salient than others (Bordalo, Gennaioli and Shleifer, 2012, 2013), they will naturally attract more attention, resulting in reduced error rates.

However, studying the dependence of error rates on underlying economic variables is far from straightforward, for two reasons which we will elaborate on below. The first is that it is by no means clear what the gradual effects predicted by psychology and neuroscience should depend on for *economic* choices, where a natural scale as weight, brightness, or length is usually not part of the problem's formulation, and utilities are neither directly observable nor objective. The second is that, even if one glosses over the former point, and although gradual effects transforming utility differences in economic choices are often assumed in applied economics (Anderson, Thisse and De Palma, 1992; McFadden, 2001; Moffatt, 2015), the estimation method might often create apparent regularities where none exists, hence obscuring the actual origin of the key regularities.

The first problem is an obvious one. In a sense, and with apologies to those fields, psychophysicists and perceptual psychologists face easier problems than economists. It is *a priori* not clear whether objectively-given scales might play the role of weight or length for economic decisions, or even for some of them. For instance, on the basis of the available evidence, a good case could be made for numerical magnitudes, independently of whether they are payoff-relevant or not. Results by Moyer and Landauer (1967) and Dehaene, Dupoux and Mehler (1990) (see also Dehaene, 1992; Dehaene et al., 2008) show that the gradual effects on error rates and response times exist even when humans are asked to discriminate among single-digit numbers. That is (astonishingly), people make more mistakes (and take longer to decide) when asked whether 6 is larger than 5 than when asked whether 9 is larger than 2. This is compatible with evidence from electroencephalography (EEG), which suggests that the neural representations of numbers vary in a continuous, gradual way with numerical distance (Spitzer, Waschke and Summerfield, 2017). Recently, Khaw, Li and Woodford (2018) have suggested that the mere imprecise representation of numerical magnitudes along these lines may explain the large estimates of risk aversion which are typically observed in laboratory experiments in economics. Given that many economic tasks come with a numerical framing, it is necessary to tackle the question of which is the (most) relevant dimension underlying possible gradual effects on economic choices.

Although error rates might indeed be affected by the imprecise representation of observational variables as e.g. numerical magnitudes, this is unlikely to be the only determinant of errors. More natural candidates, painting a less bleak picture of human rationality, are related to economic gains, as is the case of expected value or (estimated) expected utility. Inspired by models from psychology (Thurstone, 1927), random utility models, as pioneered by Marschak (1960) and McFadden (2001), assume that errors depend on underlying (unobservable) utility differences. However, very few studies have actually empirically demonstrated a monotonic relation between error rates and differences in underlying utilities. A notable exception is the early study of Mosteller and Nogee (1951), which used utilities estimated through an interpolation procedure. An added difficulty is that both in the empirical work of Mosteller and Nogee (1951) (and other experiments), as well as in theoretical random utility models an error is *defined* as a choice which does not maximize utility, that is, there is no *ex ante* definition of error independent of the (estimated or assumed) utility.

This leads us straight to the second problem. At least since McFadden (2001), most applied work in discrete choice microeconomics *assumes* a gradual relation between underlying utility differences and choice probabilities, often with a specific logit or probit form, in order to parametrically estimate the utilities themselves. While this approach is invaluable to compare the fit of different utility-based models of choice and has delivered important insights, it is not appropriate to test the basic hypothesis that gradual effects exist, or to pin down measurable determinants thereof. To drive this point home, we constructed a dataset by simulating fictitious subjects who made completely random decisions among alternative risky choices. We then treated the dataset as if it would come from actual decision makers and used a standard fitting approach estimating an alleged risk propensity, assuming that errors depend on utility differences. Specifically, we assumed a CARA utility function and heteroskedastic errors with a logit form, as commonly done in the literature (e.g., Moffatt, 2015); for more details on the estimation procedures, see Section 4.2 below. Plotting choice frequencies against the estimated utility differences yields a regular sigmoidal curve (as in any logit or probit model), which creates the appearance of order (and gradual effects arising from utility differences) for the nonsensical dataset. This is shown on the left-hand panel of Figure 1. Actually, this appearance is a mere artifice of the method, as can be shown by estimating utility *out of sample*, i.e., using part of the choices for estimation purposes and plotting the rest of the choices against the resulting estimated utility differences. Specifically, we estimated individual risk attitudes using a random parameter model (Loomes and Sugden, 1995, 1998), which in particular does not assume a logit form for error terms (again, see section 4 for details on the estimation). We used even-number choices to estimate a utility which we used to plot data from odd-numbered choices, and vice versa. This approach shows

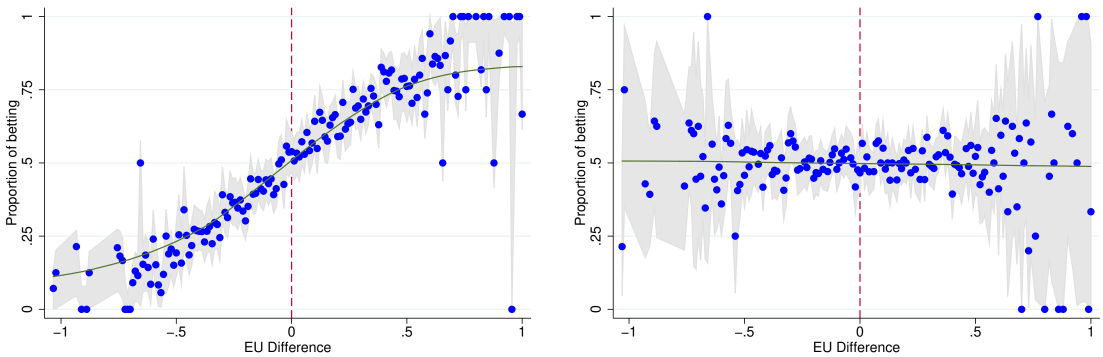

Figure 1: Analysis of a dataset of random, simulated choices. Left-hand panel: Choices as function of expected utility difference using a standard, within procedure. An appearance of order and gradual effects of expected utility differences on error rates emerges, even though no regularity is present in the data. Right-hand panel: The same choices as function of expected utility difference using an out-of-sample procedure. No regularity can be identified. Gray areas indicate 95% binomial proportion confidence intervals.

that there is no actual regularity in the dataset, as depicted on the right-hand panel of Figure 1. We conclude that structural models where utility is estimated can mistakenly create an appearance of gradual effects, and hence direct tests are needed.

In this work, we aim to test and clarify the dependence of error rates (the variable of interest) in decisions under risk on economic variables. We attack the problem on three fronts. First, we conduct an experiment (Experiment 1) with normatively-correct answers (but high error rates) where the explanatory variable can be determined in advance. This is made possible by employing a binary-choice gambling task where the winning probability or, equivalently, expected value, is an unequivocal, objectively measurable indicator of choice difficulty. In this study, the definition of error can be made *ex ante*, independently of any estimation of utility, simply because correct responses are independent of attitudes toward risk. In this way, we commit to the explanatory variable *before collecting the data*, and utility estimation plays no role. The task is simple in the sense that rational decision makers could "figure it out" with relative ease, as is the case of many problems in judgment and decision making, where an optimal decision under risk has to be made on the basis of individual beliefs. We find a significant fraction of errors (above 25%), and we demonstrate that far from being pure noise, error rates stand in a clear monotonic relation with differences in expected value. That is, we demonstrate the existence of gradual effects of an *objective economic distance* among alternatives and error rates. The design also allows us to test for dependence on payoff-irrelevant numerical effects as in Moyer and Landauer (1967), and we find that there is indeed some relation, but it is a second-order phenomenon compared with the dependence on expected-value differences. This delivers a first, objective confirmation of psychophysical, gradual effects in decision making under risk arising from economic variables.

Second, we conduct a different experiment (Experiment 2) using a betting paradigm where whether a decision is correct or not depends on individual attitudes toward risk, as is the case for most lottery-choice tasks. Rather than fitting the data to an estimated utility, we employ an out-of-sample estimation procedure excluding any artifices arising from the estimation method. Again, we find a monotonic relation, with larger error rates arising when the differences in the expected utility of the options are smaller. In contrast, the dependence on expected value differences is considerably weaker. That is, we demonstrate that the gradual effects on choice observed in psychophysics can be readily found in standard economic tasks, but they will in general arise from a *subjective economic distance* which arises from integrated, unobservable variables ("utility"). Further, economic distance (subjective or objective) can then be considered a *cardinal* measure of "strength of preference," because its cardinal magnitude determines a measurable, continuous variable (error rates).

Third, we conduct an additional, confirmatory test. In both experiments, we collect data from response times as an independent variable, which in particular plays no role for the estimation of underlying utilities. Psychophysics predicts a robust relation, with decisions where stimuli are closer being slower. We find this relation in both experiments. In Experiment 1, response times increase as differences in expected value decrease, but they are relatively unaffected by payoff-irrelevant numerical magnitudes. In Experiment 2, the analysis of response times confirms that differences in underlying utilities are a better candidate for economic distance (which replaces the choice difficulty of psychophysics) than differences in expected value (or numerical magnitudes). In both cases, the relation with response times (again a measurable, continuous variable) confirms the cardinal content of economic distance.

Taken together, our evidence demonstrates that the psychophysical effects found in the cognitive sciences are indeed very relevant for economic decisions under risk, but they depend more on economic variables than on perceptual or numerical ones. Decision-irrelevant factors (numerical magnitudes) influence error rates, but they play a secondary role in comparison with purely-economic variables. In settings where objectively-optimal answers can be derived from (correct) beliefs, it is possible to give an exogenous definition of errors, which in turn allows for a straightforward observation of the link between economic distance and error rates (in particular, one which is free of estimation problems). When risk attitudes play a role, the explanatory variable is a subjective, integrated one capturing "strength of preference," which needs to be estimated. This result validates the ideas and assumptions behind random utility models. Further, the relation to response times shows that the effects are more than "as if" accounts of decision making and have their origin in brain processes of a gradual nature, as assumed e.g. by evidence accumulation models (Ratcliff, 1978; Fudenberg, Strack and Strzalecki, 2018).

The paper is structured as follows. Section 2 briefly reviews the related literature. Sections 3 and 4 discuss Experiments 1 and 2, respectively. The analysis of response times is conducted in the last subsections within those sections. Section 5 concludes.

2 Related Literature

Our work is related to long-standing problems in economics and to several strands of the recent literature in economics. The study of stochastic choice and random utility models, going back to classic contributions as those of Debreu (1958) and Luce (1959), has endorsed the view that utilities should be understood as reflecting choice probabilities, in direct opposition with the neoclassical view that they reflect preferences of an exclusively ordinal nature (Hicks and Allen, 1934). The proliferation of experimental data showing the stochastic nature of economic choice has led to increased attention on theoretical models of stochastic choice in the recent years (e.g. Manzini and Mariotti, 2014; Matějka and McKay, 2015; Fudenberg and Strzalecki, 2015; Apesteguía, Ballester and Lu, 2017; Apesteguía and Ballester, 2018). In game theory, models of stochastic (logit) choice based on observable payoffs and unobservable idiosyncratic shocks have given rise to new equilibrium concepts as quantal response equilibria (McKelvey and Palfrey, 1995; Goeree, Holt and Palfrey, 2005). In microeconometrics, models of discrete choice (Anderson, Thisse and De Palma, 1992) have become standard for fitting experimental data and recovering underlying utility functions, frequently under "Fechnerian" assumptions (Fechner, 1860) which postulate a logit or probit form for error terms (see Moffatt, 2015, for a detailed overview). Those models assume exact functional forms mapping differences in utilities to error terms, which are highly valuable as structural assumptions but are in general not directly tested. In stark contrast, Alós-Ferrer, Fehr and Netzer (2018) have recently shown that certain properties of the empirical distribution of response times allow to recover the underlying preferences in random utility models without imposing any substantive assumptions on the distribution of random terms.

To the best of our knowledge, the first study to point at a connection between utility differences and choice frequencies was the inspiring experiment of Mosteller and Nogee (1951) on poker dice gaming, which aimed to "test the validity of the construct" represented by (expected) utility. Their analysis included illustrations which suggested a sigmoidal relation between utility differences and choice frequencies, although, as the authors admitted, those were at the individual level and cherry-picked among all experimental participants. While suggestive, their illustrations were not a test for the presence of gradual effects (and were actually not meant to be), because their utility functions were constructed exclusively out of observed indifferences. For instance, although their illustrations map zero utility difference to 50 percent choice frequency, "this finding was built into the expected utilities by the construction leading to the utility curves" (Mosteller and Nogee, 1951, p. 202).

Conceptually, our work is also related to the study of Khaw, Li and Woodford (2018), who carried out an experiment on risky choice where participants chose between a sure amount and lotteries with a single non-zero outcome and a fixed probability of winning varying amounts (that is, the winning probability was identical for all choices). By varying the sure amount and the lottery outcome, Khaw, Li and Woodford (2018) explored the reaction of choice frequencies to changes in payoffs and argued that the data could be explained assuming an imprecise internal representation of numerical magnitudes, in line with Moyer and Landauer (1967) and Dehaene (1992). Hence, their work speaks in favor of a direct effect of numerical magnitudes in error rates. However, by design, their numerical magnitudes stand in a monotonic relation to payoffs, and hence in their data it is not possible to disentangle the effects of numerical magnitudes and the effects of expected values (or utilities). In our experiments, different numerical magnitudes are associated with the same expected payoffs and vice versa, allowing us to study the effects separately.

The cardinal effects derived from a strength-of-preference account might also be helpful to improve our understanding of standard behavioral phenomena. One example is the asymmetric dominance or "decoy" effect, where the addition of a dominated option shifts the choice frequencies in a previous pair in favor of the option dominating the added one. Sürücü, Djawadi and Recker (2019) point out that this effect might decrease with the strength of preference among the two original options, i.e. might be strong enough to overturn a weak preference but not a relatively strong one. Soltani, De Martino and Camerer (2012) find within-subject decoy effects to be increasing with the distance between the decoy and the original options, derived as a difference in estimated utilities. While these contributions focus on the decoy effect and do not provide a direct test of the gradual effects we are interest in, they do show that these effects are both plausible and consequential in economic contexts.

Although our main variable of interest are decision errors, we also examine response times for two reasons. First, well-established effects in psychophysics encompass both error rates and response times, hence the analogy would not be complete without the latter. Second, while an explanation of error rates alone might be challenged as a pure "as if" story, response times allow reasonable inferences on the actual decision processes generating the errors. In this sense, our work is related to the small but growing literature examining response times in economics (see Spiliopoulos and Ortmann, 2018, for a recent review). Chabris et al. (2009) studied intertemporal decisions and found a monotonic relationship between response times and estimated utility differences (discount factors). Alós-Ferrer et al. (2016) postulated a model of lottery choice and evaluation including a relation between choice difficulty and response times to investigate the determinants of the preference reversal phenomenon (Lichtenstein and Slovic, 1971; Grether and Plott, 1979; Tversky, Slovic and Kahneman, 1990). Other response-time studies in economics include Wilcox (1993, 1994) and the web-based studies of Rubinstein (2007, 2013). On a different front, Achtziger and Alós-Ferrer (2014) relied on response times to differentiate different decision processes in a framework where intuitive reinforcement might conflict with optimal decisions based on Bayesian updating of beliefs (see also Alós-Ferrer and Ritschel, 2018).

3 Experiment 1: Objective Domain

We first aim to demonstrate the gradual effects of "strength of preference" on error rates in a domain where the variable influencing those effects is objectively given, and, as a consequence, utility estimation plays no role. The task we employ is representative of studies in the judgment and decision-making domain, where economic agents make decisions under risk or uncertainty but there is an objectively-correct answer, for example due to stochastic dominance. A prominent example is given by tasks involving updating of previously-held beliefs (e.g. Grether, 1980, 1992; Charness and Levin, 2005; Achtziger and Alós-Ferrer, 2014). We will rely on a simple gambling task with given probabilities, which is designed with two objectives in mind. The first is that objectively-correct decisions exist, independently of attitudes toward risk, and thus an exogenously-given measure of the strength of preference is available. The second is that numerical differences (in a perceptual sense) can be disentangled from economic incentives, allowing us to investigate both possible dimensions of choice difficulty.

3.1 Design and Procedures

The experiment was computerized and programmed in Psychopy (Peirce, 2007), a software which ensures high precision in the measurement of response times. We recruited $N = 96$ participants (54 females, age range $19 - 47$, mean 24) using ORSEE (Greiner, 2015) at the Cologne Laboratory for Economic Research. Participants were university students enrolled in fields other than psychology and economics. They were provided with written instructions and answered five control questions before starting the task, to ensure correct comprehension of the procedures and payment mechanism. Three participants were unable to understand the task and were excluded from the analysis. Subjects were paid according to their performance in the experiment. Total earnings were the sum of the earnings in the 160 trials plus a show-up fee of EUR 4. Sessions lasted around 60 minutes and the average payoff was EUR 16.45 (around USD 17.60 at the time of the experiment).

The experimental task is as follows. Participants are confronted with three decks of cards, a red one (Diamonds) and two black ones (Clubs and Spades), containing ten cards each (numbered 1 to 10, see Figure 2). The participants' task is to choose *twice* from which of the two black decks a card should be randomly extracted from, and the game's objective is to beat a card extracted from the red deck with the black one. Each trial starts with a participant choosing between the two (complete) black decks, but this first choice is irrelevant for our purposes since at this point both decks are identical. It is also unpaid (to avoid possible reinforcement or valence effects). The choice, however, creates an asymmetry which is the essence of the task. After the first choice is made and the first black card is extracted, that card remains on the table (there is no replacement). A card is extracted from the red deck, and the participant is asked to choose between the black decks a second time. This is the choice we are interested in. A (black) card

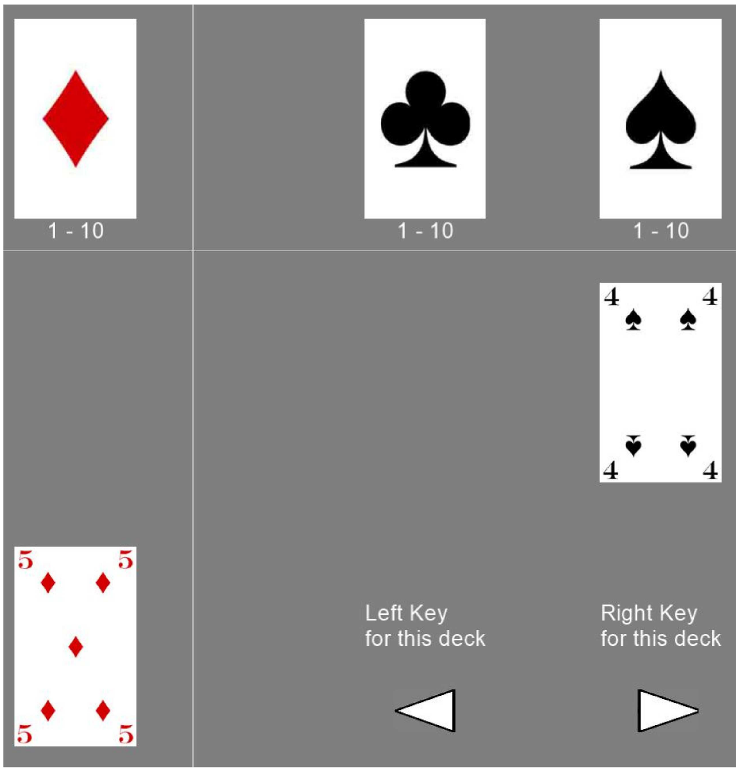

Figure 2: Experiment 1. A trial starts with participants choosing between the black decks. Consequently a black card is extracted from that deck and a red card is displayed. No replacement happens after the first choice. Participants then choose again between the black decks, and a card is extracted from the chosen deck. If the second extracted black card is strictly larger than the red card, the participant wins, otherwise she loses.

is extracted from the chosen deck, and the participant received EUR 0.15 if and only if that black card has a number strictly bigger than the red card, otherwise she receives nothing. Subsequently, the trial ends, all cards are placed back in their decks, and decks are reshuffled before a new repetition starts. Participants knew that trials were independent, so the outcome as well as the cards displayed in one trial were unrelated to those of subsequent trials. Each participant completed 160 of such trials.

After the first choice, one black deck has either one winning card less or one losing card less. Hence, by design, there is an optimal decision pattern for the second choice, which is to bet on the deck with a higher proportion of winning cards. That is, if the first black card was smaller or equal than the red card, the participant should choose the same deck, and if the first black card was strictly larger than the red card, the participant should choose the other deck. In the example depicted in Figure 2, the red card is a 5 and the first black card is a 4 (of spades), so the spades deck contains only 9 cards, 4 losing and 5 winning ones. The clubs deck still contains 10 cards, 5 losing and 5 winning cards. Therefore the deck of spades contains 1 losing card less than the untouched deck of clubs and the optimal choice is to choose it again. On the contrary, if the first extracted black card had been strictly larger than the red card, the chosen deck contains 1 winning card less than the other one, and the optimal decision would be to choose the untouched deck. Hence, independently of risk aversion, there is always a normatively-correct decision for the second choice.

In spite of the fact that all choices are either objectively correct or objectively wrong, some choices are "more correct" than others, because opportunity costs are different. Let $r, b_1 \in \{1, \ldots, 10\}$ be the red card and the first extracted black card, respectively. Let $\pi^0(r, b_1)$ and $\pi^1(r, b_1)$ be the probability of winning by choosing the same deck or by shifting to the other deck, respectively. Then $V(r, b_1) = \left|\pi^0(r, b_1) - \pi^1(r, b_1)\right|$ is the cardinal difference (distance) between the probability of winning by making the correct choice and the probability of winning while making an error. Since participants are paid only in case they win, up to a rescaling of monetary units this is also the difference in expected values between a correct decision and an error. If $r \geq b_1$, one obtains $\pi^0(r, b_1) = (10 - r)/9$ and $\pi^1(r, b_1) = (10 - r)/10$. If $r < b_1$, one obtains $\pi^0(r, b_1) = (10 - r - 1)/9$ and $\pi^1(r, b_1) = (10 - r)/10$. Hence,

$$V(r, b_1) = \begin{cases} (10 - r)/90 & \text{if } r \geq b_1, \\ r/90 & \text{if } r < b_1. \end{cases}$$

By design, $V(r, b_1)$ assumes values in the set $\{1/90, 2/90, ..., 9/90\}$. These differences in expected value indicate the opportunity cost of (not) choosing the right answer and reflect how far away from "indifference" the participants were in every decision, and are hence a natural measure for the "strength of preference." Thus, we take $V(r, b_1)$ as the potential driver for stochastic choice and refer to this magnitude as *(objective) economic distance.*

The probabilities of winning by staying or switching, and the economic distance, are monotonic functions of the numerical value of the red card. However, for computing the optimal choice the only necessary information is the sign of the relation between the first black card and the red card. That is, the actual magnitude of the difference between the values of these two cards is economically inconsequential. However, Moyer and Landauer (1967), Dehaene, Dupoux and Mehler (1990), and others have shown that, in simple comparisons, errors do depend on the numerical differences between stimuli. Therefore, we also contemplate the possibility that the distance between the numerical values of the first black card and the red one influences choice frequencies (and response times). There are ten possible distances between the two cards, ranging from 0 to 9. We refer to this magnitude as the *numerical distance.*

To ensure enough variability in the stimuli, the set of initial stimuli (first black card and red card) was predetermined and pseudorandomized. Furthermore, the red card was never a 10, since in this case winning would be impossible, hence the choice would be inconsequential (the instructions did not claim that the red card was randomly selected, since the procedure by which it was selected was payoff-irrelevant once the actual choice was faced). The key second black card was randomly selected among the remaining cards.

3.2 Choices and Errors

In spite of the simplicity of the task, the mean error rate across participants was 28.93%, with a median of 31.25% (SD = 18.21, min 0.63%, max 60.00%). We start by examining the dependence of error rates on both economic and numerical distance. Figure 3 plots the frequency of "stay" decisions (choosing the same deck as in the first decision) for each possible value of each variable. The left-hand side panel plots the dependence on economic distance, i.e. differences in expected values. The red shaded areas correspond to errors, as a rational decision maker should stay for a positive expected value difference and switch for a negative one. To facilitate the comparison, in all figures and regressions the economic and numerical distances are both normalized to be between 0 and 1.[3] Clearly, the probability to stay with the same deck stands in a clear positive relation with the difference in expected values. The frequency of errors becomes smaller as the difference becomes larger (no matter the sign), and it is largest (essentially 1/2) when the difference approaches zero. This pattern is radically different from that predicted by neoclassical economic theory. Even accounting for noise, neoclassical predictions would prescribe choice frequencies with a flat slope somewhere above zero for negative values of the expected value difference (where stay is an error), and a flat slope somewhere below one for positive values (for which stay is the correct option). This is clearly not the case. Subjects gradually make less errors as the objective economic distance between the options becomes larger.

In contrast, the right-hand panel depicts the relation between choice frequency and the numerical difference between the values of the (first) black card and the red card. Again, shaded areas correspond to errors, since the correct decision is to stay when the first black card was strictly smaller than the red one, and switch otherwise. The gradual relation is essentially absent in this case (with slopes being relatively flat), and there is a clear discontinuity at zero (the normative switching point). That is, subjects on average understood the task, but the numerical distance between card values does not appear to play a large role. The comparison between the panels suggests that the variability in responses arises mainly from differences in expected values, and not from purely numerical differences.

Figure 4 further investigates the relative contribution of the two dimensions of choice difficulty by letting one variable vary while keeping the other fixed (which is made possible by our design). In the left-hand panel, we plot choice frequencies as a function of expected value differences, separately for trials where the numerical distances corre-

[3] For all figures, unless otherwise specified, each point represents each distinct value of the variable in the x-axis and the corresponding average value of the variable in the y-axis (choice frequencies or average response times). Therefore each point is an average across potentially different subjects and trials. The depicted curves are estimated using a fractional regression with a polynomial of second degree.

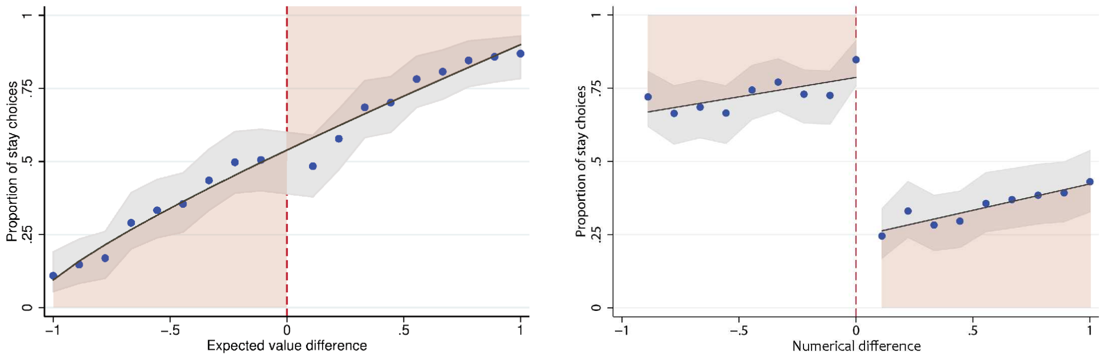

Figure 3: Experiment 1. Stay choices as function of expected value difference between staying and switching the deck (left-hand panel), and as function of the numerical distance (black minus red; right-hand panel). Gray areas indicate 95% binomial proportion confidence intervals.

spond to three particular, fixed values (1, 2, or 3, corresponding to 0.1, 0.4, and 0.7 after normalization). The positive relation between the proportion of stay choices and expected value difference is essentially unchanged, with the depicted curves essentially overlapping. In contrast, the right-hand panel plots choice frequencies as a function of numerical distances, separately for trials where the expected value differences correspond to three particular, fixed values (1/90, 2/90, or 3/90, again corresponding to 0.1, 0.4, and 0.7 after normalization). The relation changes drastically for different expected value differences, uncovering a negative, monotonic, and gradual relation between the proportion of stay choices and the numerical difference between the stimuli which becomes stronger for larger expected value differences. The figure suggests again that expected value differences are the determinant factor (gradually) influencing error rates, but also that, when keeping the economic dimension of choice difficulty fixed, second-order effects appear which are compatible with common findings from the perceptual literature.

We now turn to a regression analysis. The data form a strongly balanced panel with 160 trials for each of the 93 participants. Table 1 shows random-effects Probit regressions where the dependent variable is 1 in case of a correct answer.[4] Model 1 establishes the basic effect, namely that larger (objective) economic distances lead to less errors. Model 2 introduces the numerical distance and shows that this variable also leads to lower error rates, revealing perceptual effects on top of value-induced ones.[5] Since both distances are normalized, we can compare the magnitude of the two effects. The regression coefficients for economic distance range from 1.2 to 1.6, and the regression coefficients for numerical distance range from 0.36 to 0.58. We can also calculate the relative elasticity of the two variables. A percentage variation in the economic distance predicts an average increase

[4]Using a fixed effects regression instead does not affect the main results, showing that these results are not determined by heterogeneity among subjects. The same comment applies to all other panel regressions below.

[5]The results are unchanged if we define the numerical distance through the log of the numerical values of the cards, as suggested by Moyer and Landauer (1967).

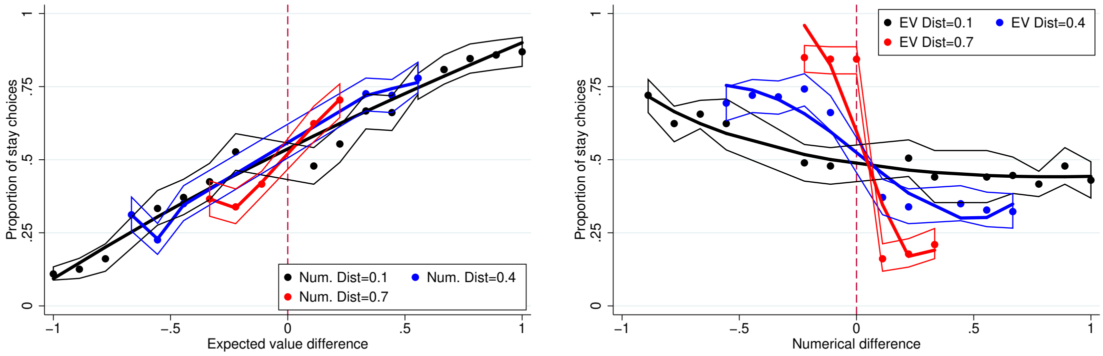

Figure 4: Experiment 1. Left-hand panel: Stay choices as a function of expected value differences for various, fixed numerical distances. Right-hand panel: Stay choices as a function of numerical distance for various, fixed expected value differences. The 95% binomial proportion confidence intervals are plotted.

of 21.84% in the probability of a correct answer, while a percentage variation in the numerical distance predicts an average increase of only 5.40% in the probability of a correct answer. This provides further evidence for the predominant role of the economic dimension of choice difficulty over the perceptual.

As Figure 4 illustrates, our design allows to examine trials with identical expected value differences but different numerical distances, and vice versa. However, a purely mechanical effect prevents both variables from being fully orthogonal, as a larger numerical distance between the cards allows a larger number of feasible values of economic distance (the Spearman correlation between numerical and economic distances across the set of decisions is $\rho = -0.6491$, $N = 160$, $p < 0.001$). Hence, in Model 3, we introduce the interaction between the two dimensions of choice difficulty as a control. The coefficient is significant and negative, reflecting the mechanical relation in the dimensions across the entire dataset. However, the main effects are unaffected by this control.

Last, Model 4 adds a number of other controls: gender, native language, left-handedness, and cumulated earnings (Sum Won). The regression shows that the main effects are robust. Females (54) make more errors than males, as can be confirmed by a direct, non-parametric test (females 33.08%, males 23.19%; Mann-Whitney-Wilcoxon test, $N = 93$, $z = 2.764$, $p = 0.0068$). Native speakers (72) make less errors than other participants (natives 25.89%, others 39.35; MWW test, $N = 93$, $z = 3.116$, $p = 0.0018$). Also, participants who earned more in previous trials are more likely to make a correct choice, which is merely an indication of heterogeneous skills among participants. In all regression models, we control also for learning effects (round, 1 to 160) and find that the probability of making an error decreases over time.

Table 1: Experiment 1. Random-effects Probit regressions on correct answers.

Correct	Model 1	Model 2	Model 3	Model 4
Econ. Dist.	1.229***	1.447***	1.622***	1.620***
	(0.099)	(0.107)	(0.140)	(0.141)
Num. Dist.		0.362***	0.584***	0.582***
		(0.087)	(0.094)	(0.093)
Econ. Dist. × Num. Dist.			−0.010***	−0.010***
			(0.004)	(0.004)
Sum Won				0.011*
				(0.006)
Female				−0.438***
				(0.141)
Native				0.462***
				(0.169)
Age				−0.045***
				(0.010)
Left handed				0.378**
				(0.194)
Round	−0.001***	−0.001**	−0.001**	−0.007**
	(0.000)	(0.000)	(0.000)	(0.003)
Constant	0.233**	0.003	−0.061	0.886***
	(0.095)	(0.096)	(0.103)	(0.284)
Log L.	−7358.011	−7337.550	−7329.264	−7265.955
Wald test	163.304***	191.995***	188.653***	251.391***
Obs.	14880	14880	14880	14880

Robust standard errors in brackets, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

3.3 Response Times and the Underlying Processes

As an additional, independent test of the cardinal effects of economic distance, we measured response times for all decisions. The key response time for our purposes is the one of the second decision within each trial. Other response times, however, can be used to control for individual differences in (mechanical) swiftness, e.g. as arising from the relative ease of interface use. For the second decision, we computed individual average response times. The average of those across individuals was 1.612 seconds (SD = 0.648, median 1.560, min = 0.272, max = 3.797).

The left-hand panel of Figure 5 plots average response times as a function of the expected value difference between stay and switch. There is clear evidence of gradual effects as postulated in psychophysics. An inverted U-shape is apparent, indicating a negative relation between response times and the distance in expected values between stay and switch.[6] Choices closer to indifference (zero expected value difference) are as-

[6] A random-effects regression verifies the significance of the curvature. The coefficient of the squared expected value difference is significantly negative (Coef. $-0.169, z = -7.97, p < 0.001$).

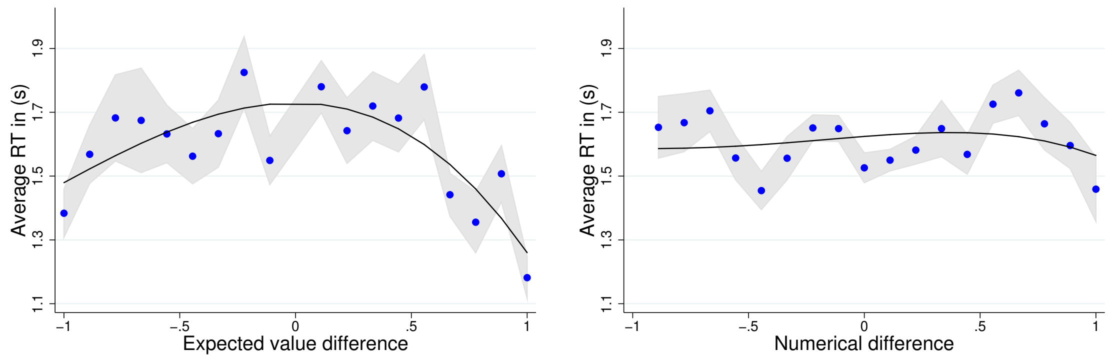

Figure 5: Average response times as a function of expected value difference (left-hand panel) and as a function of numerical distance (right-hand panel).

sociated with the longest response times. In contrast, the right-hand panel of Figure 5, which depicts the relation between response times and numerical differences, shows an essentially flat trend. That is, unlike in the case of expected value differences, there is no discernible pattern. In summary, response times suggest a gradual effect of (objective) economic distance (but not of numerical distance), confirming that the postulated relationship goes beyond a simple *as if* story and reflects actual decision processes.

We now turn to a regression analysis. Response times are a noisy variable, usually presenting a skewed, non-normal distribution and rare extreme observations. To account for these features it is common practice to take the logarithm of response times as the variable of interest in regression analyses (Fischbacher, Hertwig and Bruhin, 2013; Achtziger and Alós-Ferrer, 2014). Table 2 reports random-effects regressions of log-transformed response times, taking advantage of the panel structure of the data. To control for individual differences in mechanical swiftness, we use the log-transformed response time for the non-rewarded, first black card (RT1) and the log-transformed response time for pressing a space bar, which was required before the start of each trial (RT0). Model 1 establishes the basic effect, namely that responses are faster for larger (objective) economic distances, confirming that the phenomena we study reflect basic properties of actual decision processes. Model 2 adds numerical distance. The coefficient is also significantly negative, although of a smaller magnitude (recall that both variables are normalized to have the same range). This shows that, in spite of the relatively flat shape of the aggregate relation as depicted in Figure 5, response times are also influenced by numerical differences, at least as a second-order determinant. As in the case of choice frequencies, Model 3 shows that the effects are robust to controlling for the interaction between the two distances. Finally, Model 4 shows that the results are robust to additional controls. Gender and cumulated earnings did not affect response times, but native speakers took longer to respond than other participants.

Table 2: Experiment 1. Random-effects regressions on log response times.

Log RT	Model 1	Model 2	Model 3	Model 4
Econ. Dist.	-0.174^{***}	-0.237^{***}	-0.321^{***}	-0.316^{***}
	(0.020)	(0.031)	(0.035)	(0.035)
Num. Dist.		-0.106^{***}	-0.224^{***}	-0.209^{***}
		(0.029)	(0.038)	(0.039)
Econ. Dist. × Num. Dist.			0.005^{***}	0.005^{***}
			(0.001)	(0.001)
Sum Won				0.008
				(0.006)
Female				-0.056
				(0.078)
Native				0.281^{***}
				(0.119)
Age				-0.014
				(0.009)
Left handed				-0.014
				(0.191)
Round	-0.001^{***}	-0.001^{***}	-0.001^{***}	-0.005^{***}
	(0.000)	(0.000)	(0.000)	(0.003)
RT0	0.167^{***}	0.167^{***}	0.169^{***}	0.117^{***}
	(0.027)	(0.027)	(0.027)	(0.037)
RT1	0.126^{***}	0.127^{***}	0.127^{***}	0.145^{***}
	(0.034)	(0.034)	(0.034)	(0.032)
Constant	0.644^{***}	0.712^{***}	0.746^{***}	0.853^{***}
	(0.048)	(0.055)	(0.057)	(0.253)
R^2 overall	0.108	0.109	0.110	0.130
Wald test	219.477^{***}	220.200^{***}	257.260^{***}	259.572^{***}
Obs.	14880	14880	14880	14880

Robust standard errors in brackets, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

3.4 Discussion of Experiment 1

The first experiment is probably as close as one can get to pure psychophysics in the economics domain. By using a gambling task with objectively correct answers, we can commit to the exact values of the explanatory variable before running the experiment; that is, we can rely on expected value differences and no utility estimation is needed. Still, the task is representative of the judgment and decision-making domain and remains intrinsically interesting for economic decision making. We find a robust gradual relation between cardinal, objective economic distance, as captured by expected value differences, and error rates. Error rates gradually decrease as the distance between alternatives becomes larger (decisions become easier). We also find that purely numerical effects (by how much a number is larger than another one, even if the comparison is payoff-irrelevant) do influence error rates as predicted by psychophysical studies, but this is a second-order effect and the main explanatory variable remains economic distance.

Response times confirm the gradual relationship. As predicted by psychophysics, easier decisions (in the sense reflected by objective economic distance) are faster. This is important, because response times are a direct reflection of the underlying decision processes. Hence the relationship further confirms that the gradual effects of choice difficulty do reflect actual decision processes and not just a characteristic of how the statistical model of errors fits the data.

4 Experiment 2: Subjective Domain

Experiment 1 can be seen as a streamlined proof of concept which does away with the problems inherent in utility estimation. In Experiment 2, we parsimoniously go one step further by reproducing the analysis for more complex decisions under risk where what is "correct" depends on the individual risk attitude, and hence utility estimation is unavoidable. In this sense, Experiment 2 studies choices in the subjective domain, while Experiment 1 belonged to the objective domain. Crucially, to avoid the problems pointed out in the Introduction, we will strictly adhere to an out-of-sample approach where the utility used to test the gradual dependencies in the data is always estimated from a different part of the dataset. This ensures that the estimation allows us to test for the presence of gradual effects, instead of artificially creating them.

As in Experiment 1, we focus on error rates. We will have three (explanatory) variables of interest in sight. Of course, we will focus on expected utility as just described. Additionally, in this experiment we can examine the differences between expected value and expected utility differences as determinants of gradual effects on error rates. For completeness, we will also examine the potential effects of (payoff-irrelevant) numerical distance. Last, and again as a confirmatory exercise, we will examine the effects of those variables on response times.

4.1 Design and Procedures

Implementation, procedures, and data collection were as in Experiment 1. Participants were $N = 96$ (different) university students (66 females, age range $18 - 36$, mean 24). Sessions lasted around 60 minutes and the average payoff was EUR 13.45 (around USD 14.40 at the time of the experiment). Three participants were unable to understand the task and were excluded from the analysis.

The experimental task is as follows. Participants are confronted with two decks of cards, a red one (Diamonds) and a black one (Clubs), containing ten cards each (numbered 1 to 10). At the beginning of each of the 170 trials, *two* cards are extracted from the black deck and displayed, one red card is extracted from the red deck, and a monetary prize is displayed (see Figure 6). The participants' task is to decide whether

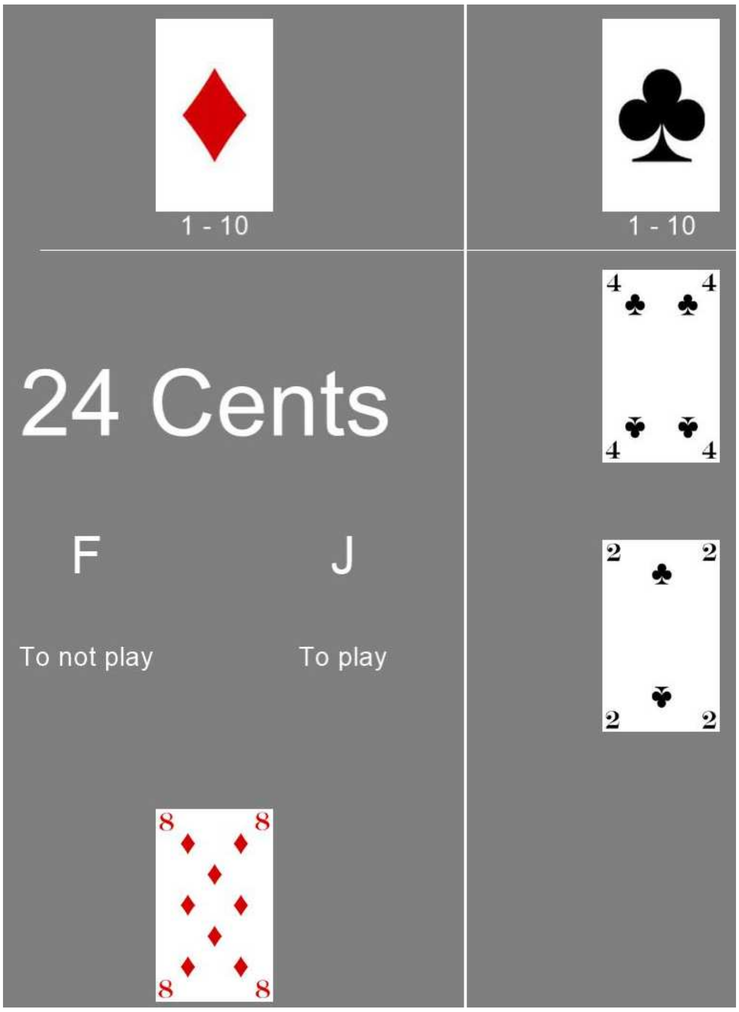

Figure 6: Experiment 2. Each trial starts by extracting two black cards, a red card, and displaying a prize. Participants then decide whether to bet or not, knowing that betting is costly. If the participant bets, a black card is extracted, and the participant wins if and only if the extracted black card is strictly larger than the red card.

to bet or to pass. After this decision, a further black card will be extracted from the remaining eight cards in the black deck, and the objective is to beat the red card with that new card. Betting is costly: placing a bet costs EUR 0.10 (fixed for all trials), independently of the outcome of the trial. If the participant bets and if the newly-extracted black card is strictly larger than the displayed red card, the participant receives the displayed monetary prize (minus the cost). Otherwise, the payment is zero (resulting in a net loss equal to the cost of betting). If the participant does not bet, there is neither a payment nor a cost, and the experiment moves to the next trial. Before a new trial starts, all cards are returned to their respective decks and those are reshuffled. Hence, each trial reflects an independent decision situation.

The set of initial stimuli (red card, first two black cards, and prize) was predetermined and pseudorandomized across trials to achieve adequate stimuli variance. The crucial third black card was randomly selected among the cards remaining in the deck. Red cards were extracted in such a way that there was always some probability of winning, so as to avoid trivial decisions. Hence, there were eight possible distinct probabilities of winning, ranging from 12.5% to a sure win. Prizes ranged from 10 to 120 cents, and were determined trial-by-trial as deviations from the actuarially-fair prize, the amount that leaves a risk-neutral agent indifferent between betting and passing. Eleven different distortions from the fair prize were implemented, ranging from 50% below to 50% above, in 10% steps.

In each trial, at the moment of the decision, the black deck contains eight cards, and the two already-extracted cards are displayed. The probability to win when betting depends on the magnitude of the red card and on whether the displayed black cards are winning or losing cards. In the example depicted in Figure 6, the red card is an 8 and the two extracted black cards are a 2 and a 4, hence both are losing cards. That is, the black deck contains two winning cards and six losing ones, yielding a probability of winning of 1/4. Since the cost of betting is 10 cents, the actuarially-fair prize is 40 cents, but the offered prize is 24 cents. Hence, a risk-averse or risk-neutral agent should decline to bet, while a risk-loving one might rationally decide to bet. That is, there are no objectively-correct decisions in this task; rather, what is "correct" depends on the individual risk attitude. Therefore, we hypothesized that the natural measure of choice difficulty or *subjective economic distance* would be the difference between the expected utilities of betting and passing, referred to as EU distance for clarity, which requires us to estimate the underlying individual utilities of money.

By design, however, the expected value of betting depends on the distortion of the fair prize. For risk neutral individuals, the difference in expected value between passing and betting reflects how far away from "indifference" the participants were, and are hence a natural, alternative measure for "strength of preference." Therefore, another candidate determinant of gradual effects is simply the absolute value of the expected value differences between betting and passing, which we refer to as EV distance. In contrast to Experiment 1, the comparison between these two measures of economic distance is informative of which is the relevant measure of strength of preference in this context.

We remark also that the probability of winning does *not* depend on the numerical distances between the black cards and the red one, but only on whether the former are larger or smaller than the latter. Hence, numerical distances in themselves are payoff-irrelevant (but the sign of the numerical differences is not). Analogously to Experiment 1, this allows us to disentangle the numerical closeness of stimuli as a further possible dimension of choice difficulty, which is the closest one to standard measures of perceptual similarity used in psychophysics. Since there are two black cards, we have different possible candidates for numerical distance. We present here the analysis using the distance between the red card and the second, most recent black card, since a large literature has advocated the prominence of the recency effect (Deese and Kaufman, 1957; Murdock Jr., 1962). We also carried out analyses with other definitions of numerical distance;

the main results described below are unaffected.[7] There are ten possible perceived distances between the red card and the second black card, ranging from 0 to 9. We refer to this magnitude as *Numerical distance*.

4.2 Utility Estimation

We estimate out-of-sample risk attitudes for each subject. Specifically, we use the choices made in odd trials to estimate risk attitudes and use this estimation to predict the expected utility in the even trials, and vice versa.[8] To derive individual risk attitudes we rely on the estimation of random parameters (Loomes and Sugden, 1995, 1998) using Maximum Simulated Likelihood (MSL) (see, e.g., Loomes, Moffatt and Sugden, 2002; Moffatt, 2005; Bellemare, Kröger and van Soest, 2008). Specifically, we adapt the estimation procedure described by Harrison (2008) and Moffatt (2015). The MSL technique is frequently used in the context of decision-making under risk (e.g., Von Gaudecker, Van Soest and Wengström, 2011; Conte, Hey and Moffatt, 2011; Wilcox, 2011; Moffatt, Sitzia and Zizzo, 2015). This approach allows us to estimate risk aversion (r) as a deterministic coefficient, but allowing for sampling error. An alternative interpretation of the procedure is that there is heterogeneity in preferences of subjects, hence r is better characterized as a distribution instead of a point estimate. For computational tractability, we assume that r follows a normal distribution in our dataset.[9]

As the functional form of the utility, we adopt a normalized CARA function as in Conte, Hey and Moffatt (2011), i.e.

$$U(x) = \begin{cases} \frac{1-\exp(-rx)}{1-\exp(-rx_{\max})}, & \text{if } r \neq 0 \\ \frac{x}{x_{\max}}, & \text{if } r = 0 \end{cases}$$

where $x_{\max}$ is the upper limit of the outcome variable x. Using a CARA utility function offers the advantage to fully accommodate zero outcomes, while at the same time it assumes away the impact of initial wealth. However, the results are robust to the use of CRRA functions.

The estimated risk propensities in our dataset have an average $\hat{\mu}_r = 0.054$ (SD $= 0.025$, median $= 0.046$, min $= 0.024$, max $= 0.163$). The risk propensity estimated on

[7]The considered alternatives were the distance between the highest black card and the red one, the distance between the average of the two black cards and the red card, and the distance between the highest or lowest black card and the red one, as well as controlling for the log transformation of the numbers.

[8]Our results do not change if we use different out-of-sample approaches, as e.g. using an initial block of observations for the estimation and predict the expected utility out of sample for the remaining trials.

[9]Classical methods based on an individual estimation of individual risk attitudes via maximum likelihood procedures avoid the distributional assumptions made by random parameter methods. However, Monte Carlo analysis shows that, with finite samples, the out-of-sample predictive performance of random utility models can be misleading when individual estimation is employed (Wilcox, 2011). Moreover, Apesteguía, Ballester and Lu (2017) have pointed out that random utility models, in the context of risk and time preferences, can violate choice monotonicity compared to natural parameterizations. Further, these works indicate that alternatives such as the random parameters methods we use, where utilities are parametric but the parameters are random, are immune to these difficulties.

odd trials ($\hat{\mu}_r = 0.054$) is not significantly different from the one estimated on even trials ($\hat{\mu}_r = 0.053$; Wilcoxon Signed-Rank test, $N = 93, z = -0.404, p = 0.6860$). The absence of negative values in both estimations shows that no subject is classified as risk-loving, while some subjects display values of r close to 0, indicating risk neutrality. However, the majority of subjects are estimated to be risk averse.

For the simulation reported in Figure 1, we generated a dataset where each of 93 fictitious subjects randomly chose 170 times between accepting certain bets or not (the dataset mimics the basic features of Experiment 2). Bets involved a certain probability of a positive prize, and led to the loss of a small amount of money with the remaining probability. The outside option always yielded zero payoffs. Prizes and probabilities changed across trials, but the amount potentially lost was fixed. The set of bets was such that a risk-neutral subject would accept half of the times. "Decisions" were fully random and unrelated to the options. The right-hand panel of Figure 1 corresponds to an estimation performed exactly as described above for Experiment 2. The left-hand panel depicts the results of an estimation using the same CARA functional form, but with a standard within-sample approach as common in the literature. Specifically, we implemented MSL assuming heteroskedastic Fechner errors (Fechner, 1860; Hey and Orme, 1994). We used the estimated risk attitudes to compute, *within sample*, the expected utility difference between the two options (betting minus passing), and then plotted this difference against the proportion of times one option was chosen over the other. As argued in the introduction, the difference between both approaches shows that the estimation procedure might create apparent gradual effects simply because they are assumed in the underlying random utility model. Our out-of-sample procedure ensures that the regularities we uncover correspond to actual features of the data.

4.3 Choices and Errors

We define an error as a choice which gives a negative expected utility, e.g. deciding to bet when the expected utility (as estimated out of sample) of betting is strictly smaller than the expected utility of passing. The mean error rate across participants was 22.71%, with a median of 20.00% (SD = 10.77, max 51.76%, min 5.29%). Figure 7 plots the frequency of betting decisions for each possible value of each variable. As in previous pictures, to facilitate the comparison, in all figures and regressions the various distances are normalized to be between 0 and 1. The upper panel plots the dependence on expected utility differences. The shaded areas correspond to errors with the definition above. We observe that the relation between betting frequency and expected utility differences has a sigmoidal shape resembling a cumulative normal distribution or a logistic curve. This shape indicates that error rates decrease gradually as the difference in expected utilities between the options becomes larger. For very large differences, error rates are close to zero. For differences close to zero, choice is essentially random (error rates close to 50%). Of course, this stands in sharp contrast to deterministic, neoclassical

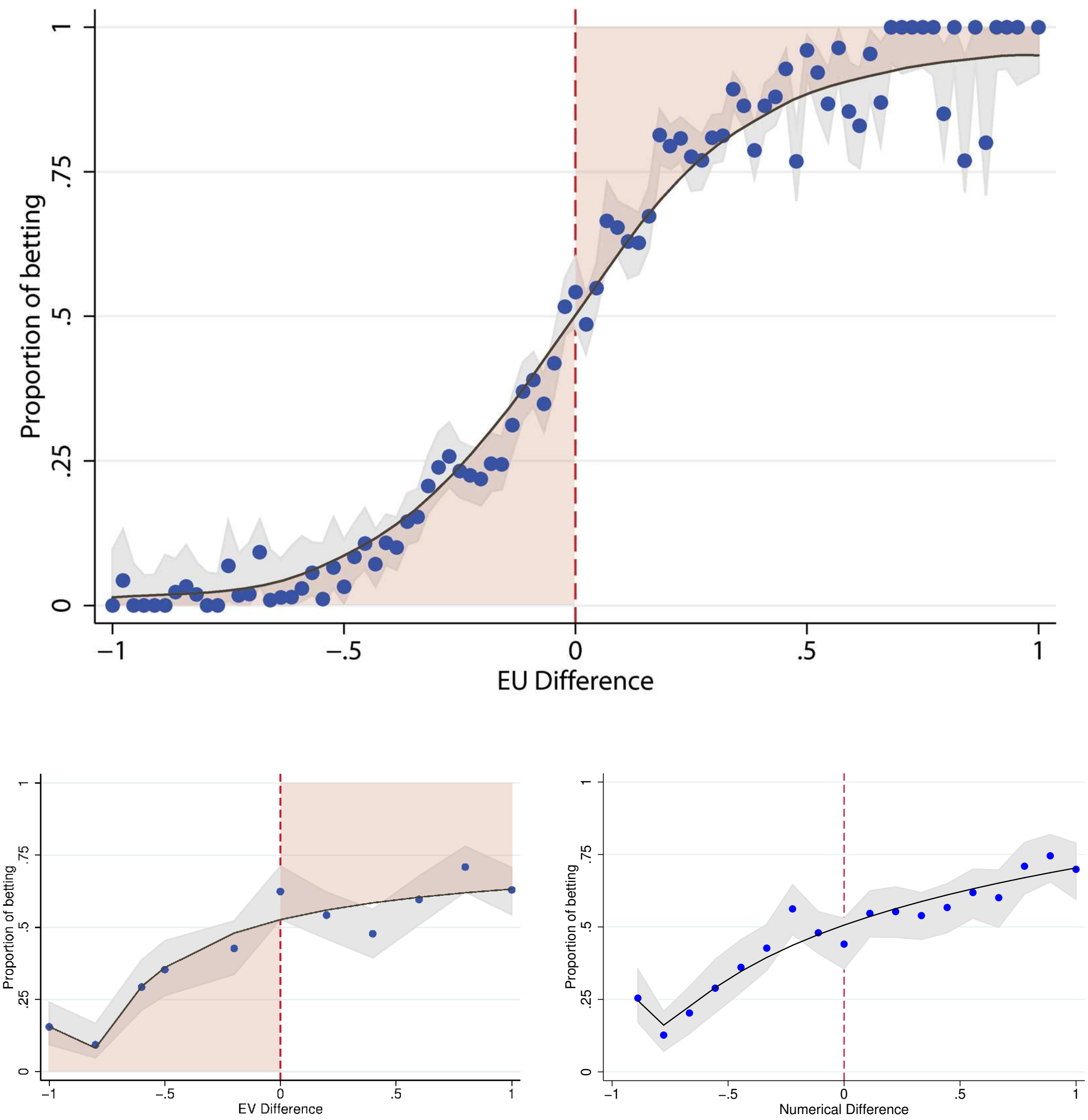

Figure 7: Experiment 2. Proportion of betting decisions as a function of expected utility differences (upper panel), expected value differences (lower left-hand panel), and numerical differences (lower right-hand panel). Gray areas indicate 95% binomial proportion confidence intervals. Shaded areas indicate the proportion of errors.

models, which would predict that subjects always bet when expected utility differences are positive and always pass when they are negative.

The lower left-hand panel plots the proportion of betting choices as a function of the differences in expected value (betting minus passing). We observe a positive but non-monotonic trend with greater expected values corresponding roughly to a higher frequency of betting.[10] This is not surprising, since as long as utility is increasing on monetary amounts, there will be some positive correlation between expected utility and expected values in a dataset. However, the figure strongly suggests that expected utility differences better explain gradual effects on error rates that differences in expected values.

Last, the lower right-hand panel plots the proportion of betting decisions as a function of numerical distances as defined above. We do not include a depiction of errors as those cannot be derived from numerical distance alone in this experiment. The picture suggests a weak, noisy monotonic relation which might hint to second-order effects but offers no strong evidence of an impact of purely numerical, payoff-irrelevant perceptions on choice frequencies. In summary, our data shows that, as in Experiment 1, there is a gradual relation between economic distance and error rates, but the former now corresponds to differences in expected utilities.

We now turn to a regression analysis. The data form a strongly balanced panel with 170 trials for each of the 93 participants. We ran random-effects panel Probit regressions where the dependent variable is 1 in case of a correct answer. For completeness, we provide separate analyses for expected utility (Table 3) and expected value differences (Table 4), while controlling for numerical distance in both. Recall that Expected Utility distance (EU distance), Expected Value distance (EV distance), and numerical distance are all normalized to range from 0 to 1. The various regression models are built in a completely analogous way, and hence we discuss them simultaneously. Note that the definitions of errors is the natural one in each table, i.e. choices which go against expected utility differences in Table 3 and choices which go against expected value differences in Table 4.

In Model 1 of both tables we see that larger economic distances lead to less errors, confirming the basic prediction. However, there is a considerable difference in the magnitude of the estimated coefficients, with EU distance having a coefficient almost 20 times bigger than EV distance. To conduct a proper comparison, we calculated the relative elasticities. A percentage variation in EU distance increases the probability of a correct answer by an average of 20.73%, while the analogous percentage for EV distance increases is only 11.98%. This confirms the message from Figure 7 that differences in expected utility, and not in expected value, are the relevant dimension of strength of preference in this context.

Model 2 in both tables introduces numerical distance as an additional control (recall the lower right-hand panel of Figure 7). In the presence of EU distance, numerical similarity between stimuli decreases the probability of a correct answer. The effect becomes marginally significant when controlling for the interaction between numerical distance and EU distance (Model 3), and loses significance when adding further controls (Model

[10]Errors in this panel are defined as decisions which contradict expected value differences. According to this risk-neutral definition, the mean error rate across participants was 36.29%, with a median of 36.47% (SD = 6.60, min 19.41%, max 54.12%).

Table 3: Experiment 2. Random-effects Probit regressions on correct answers for EU distance. Correct answer is defined as passing when $EU \leq 0$ and betting when $EU \geq 0$.

Correct	Model 1	Model 2	Model 3	Model 4
EU_Dist.	8.538***	8.836***	10.146***	10.146***
	(0.526)	(0.544)	(0.841)	(0.838)
Num._Dist.		−0.391***	−0.154*	−0.154
		(0.049)	(0.088)	(0.088)
Num._Dist. × EU_Dist.			−3.244***	−3.251***
			(1.065)	(1.065)
Sum Won				−0.000
				(0.000)
Female				−0.015
				(0.072)
Native				0.031
				(0.086)
Age				0.027***
				(0.009)
Left handed				−0.109
				(0.133)
Round	0.001	0.000	0.000	0.001
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.105**	0.249***	0.160***	−0.326
	(0.045)	(0.050)	(0.058)	(0.378)
Log L.	-7414	-7382	-7376	-7372
Wald test	281.961***	321.685***	310.504***	385.590***
Obs.	15810	15810	15810	15810

Robust standard errors in brackets,
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

4). In the presence of EV distance, numerical effects are not statistically significant. They only become significant when we further control for the interaction between numerical distance and EV distance (Model 3) as well as other controls (Model 4). The results for numerical distance should be attributed to the fact that there is a correlation between the expected value and numerical distance across all decisions in the dataset (Spearman's $\rho = 0.1453$; $N = 170$, $p = 0.0587$), but there is no correlation between numerical distance and expected utility (Spearman's $\rho = 0.098$, $N = 170$, $p = 0.2050$).

In all models we further control for learning effects. Participants appear to improve with repetition when errors are defined according to expected values, but not when they are defined according to expected utilities. There are no gender differences in errors defined according to EU distance, as confirmed by a non-parametric test (females 22.41%, males 23.46%; MWW test, $N = 93$, $z = 0.432$, $p = 0.6654$). Likewise, native speakers (77) did not perform significantly differently from other participants (natives 22.29%, others 19.41%; MWW test, $N = 93$, $z = 0.647$, $p = 0.5179$). When defining errors

Table 4: Experiment 2. Random-effects Probit regressions on correct answers for EV distance. Correct answer is defined as passing when $EV \leq 0$ and betting when $EV \geq 0$.

Correct	Model 1	Model 2	Model 3	Model 4
EV_Dist.	0.463***	0.464***	1.100***	1.113***
	(0.053)	(0.053)	(0.069)	(0.070)
Num._Dist.		−0.013	0.628***	0.639***
		(0.030)	(0.061)	(0.062)
Num._Dist × EV_Dist.			−1.552***	−1.600***
			(0.116)	(0.124)
Sum Won				−0.000
				(0.000)
Female				−0.086**
				(0.041)
Native				0.096*
				(0.054)
Age				−0.003
				(0.005)
Left handed				−0.001
				(0.070)
Round	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	−0.011	−0.006	−0.285***	−0.159
	(0.021)	(0.025)	(0.035)	(0.258)
Log L.	-10170	-10170	-10102	-10037
Wald test	223.538***	223.868***	453.437***	466.95***
Obs.	15810	15810	15810	15810

Robust standard errors in brackets,
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

according to expected values, females did behave differently (males 37.22%, females 33.91%; MWW test, $N = 93$, $z = 2.399$, $p = 0.0164$), as did native speakers (natives 35.68%, others 39.26%; MWW test, $N = 93$, $z = -1.874$, $p = 0.0609$).

4.4 Response Times and the Underlying Processes

The previous section shows that differences in expected utilities are the best candidate as an explanatory determinant of gradual effects on errors. Expected value differences and numerical differences also display significant effects, but those are of a smaller magnitude and appear less robust. In this section, we further compare the gradual effects of all three variables by focusing on response times. The main objective is to show that, while there appears to be a strong, clear correspondence between expected utility differences and actual human decision processes as reflected by response times, that relation is far from clear when it comes to other alternative variables.

The variable of interest is the time participants took to decide whether to bet or to pass. The average across individual average response times for this decision was 2.918 seconds (SD = 1.140, median = 2.687, min = 1.140, max = 7.527). Figure 8 plots average response times as a function of expected utility differences (upper panel), of expected value differences (lower left-hand panel), and numerical distances (lower right-hand panel). Response times and EU distance clearly show an inverted U-shaped relation. Harder decisions, resulting in longer response times, are those corresponding to smaller expected utility differences. However, the figure shows no systematic relation with EV differences[11] or with numerical distance. This provides an independent confirmation that a larger strength of preference, in the sense of larger subjective economic distance, can be linked to easier decisions.

As for Experiment 1, we conducted a panel regression analysis for log-transformed response times. Tables 5 and 6 report the corresponding regressions using expected utility distances and expected value distances as a measure of strength of preference, respectively. To control for individual differences in mechanical swiftness, the variable RT0 measures the log of the response time for pressing the space bar to move to the next trial.

Response times are significantly shorter for larger EU distances across all models in Table 5. This fundamental effect is robust to controlling for numerical distance, accumulated earnings, gender, native language, and other controls. Additionally, numerical distance does have an effect on response times, validating the view from psychophysics (Moyer and Landauer, 1967; Dehaene, Dupoux and Mehler, 1990) that even payoff-irrelevant perceptual differences might influence actual choice difficulty. That is, in addition to the effects of subjective economic distance, response times are shorter for more perceptually distinguishable stimuli (larger numerical distance).

In contrast, the effect of expected value differences is less clear. In Model 1 of Table 6, we observe larger response times for *higher* values of EV distance, contrary to expectations if EV distance was taken to explain the gradual effects of strength of preference. However, the effect becomes non-significant when we control for the relation between EV distance and numerical distance as well as other controls (Models 3 and 4). Again, the analysis is consistent with the view that expected utility differences are the key variable explaining the gradual effects that we investigate.

In all models we control for time trends, reproducing the standard observation that subjects become slightly faster over time. Other controls deliver no additional insights, except that left-handed subjects took longer than right-handed ones.

[11]The coefficient of the squared expected utility difference is significantly negative in a random-effects regression, coef. $= -0.090, z = -2.27, p = 0.023$. The corresponding coefficient for expected value differences is not significantly different from zero, coef. $= 0.012, z = 0.85, p = 0.393$.

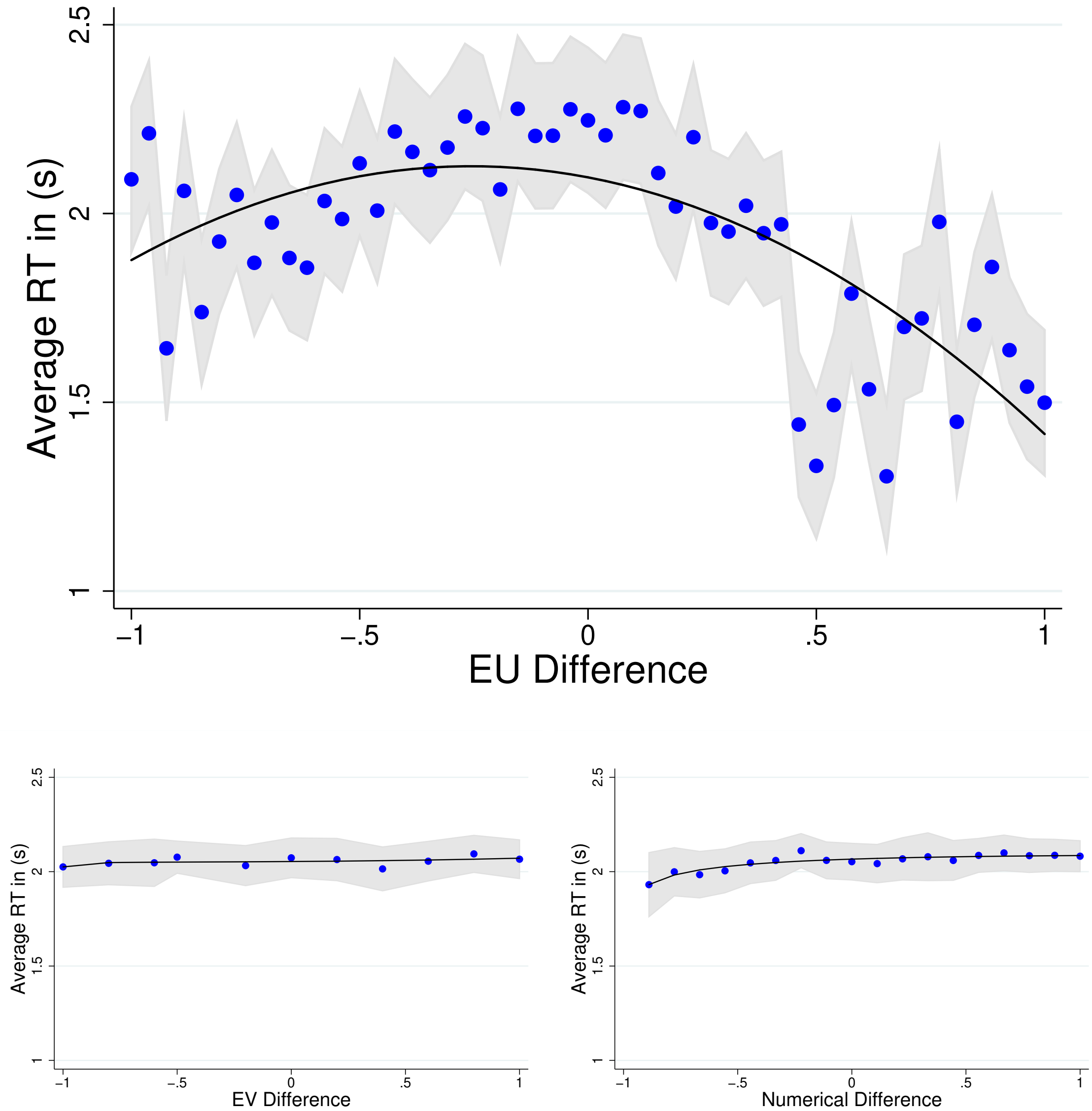

Figure 8: Experiment 2. Average response times on EU Difference (top panel), EV difference (lower left-hand panel), and numerical distance (lower right-hand panel). Gray areas indicate 95% confidence intervals.

4.5 Discussion of Experiment 2

The second experiment considers standard economic decisions under risk (betting), which are an example of preferential choice where there is no objectively correct alternative. Contrary to the first experiment, the appropriate dimension explaining gradual effects on error rates needs to be estimated from the data. Our evidence shows that expected utility, and not expected value, is the appropriate integrated variable capturing strength of preference. Choices with a larger expected utility difference between the alternatives result in lower error rates and shorter response times. The effects are robust and obtain

Table 5: Experiment 2. Random-effects regressions on log response times, EU distance.

Log RT	Model 1	Model 2	Model 3	Model 4
EU_Dist.	-1.082^{***}	-1.046^{***}	-1.255^{***}	-1.256^{***}
	(0.186)	(0.182)	(0.219)	(0.219)
Num._Dist.		-0.106^{***}	-0.158^{***}	-0.158^{***}
		(0.018)	(0.024)	(0.024)
Num._Dist × EU_Dist.			0.500^{***}	0.499^{***}
			(0.165)	(0.024)
Sum Won				0.000
				(0.000)
Female				0.024
				(0.081)
Native				0.083
				(0.114)
Age				−0.011
				(0.011)
Left handed				0.226^{***}
				(0.082)
Round	-0.002^{***}	-0.002^{***}	-0.002^{***}	-0.003^{***}
	(0.000)	(0.000)	(0.000)	(0.000)
RT0	0.199^{***}	0.201^{***}	0.201^{***}	0.201^{***}
	(0.017)	(0.017)	(0.017)	(0.018)
Constant	1.351^{***}	1.393^{***}	1.414^{***}	1.402^{***}
	(0.039)	(0.041)	(0.042)	(0.310)
R^2 overall	0.149	0.151	0.152	0.170
Wald test	562.099^{***}	686.636^{***}	708.797^{***}	735.330^{***}
Obs.	15810	15810	15810	15810

Robust standard errors in brackets,
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

even though we use a strictly out-of-sample approach, that is, they are not an artifice of the estimation method. Further, the link to response times shows that the relationship between expected utility differences and choice difficulty reflects the characteristics of actual decision processes, rather than being just "as if" modeling.

Numerical distance, seen as a more perceptual dimension of choice difficulty, plays a minor role. The effects on error rates are small and not robust to the addition of controls. Response times suggest that a second-order effect is actually present, but expected utility differences are the major determinant of the effects we study.

5 Discussion

Homo oeconomicus does not play dice (but *homo sapiens* might). A fully rational economic agent would be consistent, choosing an option 100% of the time if it delivered a slightly larger payoff than the alternative, and 0% if a minute payoff reduction left it worse than the alternative. However, considerable evidence suggests that the imple-

Table 6: Experiment 2. Random-effects regressions on log response times, EV distance.

Log RT	Model 1	Model 2	Model 3	Model 4
EV_Dist.	0.054***	0.068***	0.009	0.009
	(0.015)	(0.015)	(0.023)	(0.023)
Num._Dist		−0.149***	−0.211***	−0.211***
		(0.018)	(0.024)	(0.024)
Num._Dist. × EV_Dist.			0.145***	0.145
			(0.044)	(0.043)
Sum Won				0.000
				(0.000)
Female				0.011
				(0.083)
Native				0.073
				(0.116)
Age				−0.009
				(0.010)
Left handed				0.221**
				(0.091)
Round	−0.003***	−0.003***	−0.003***	−0.003***
	(0.000)	(0.000)	(0.000)	(0.000)
RT0	0.194***	0.197***	0.196***	0.196***
	(0.018)	(0.018)	(0.018)	(0.018)
Constant	1.231***	1.288***	1.315***	1.290***
	(0.037)	(0.039)	(0.040)	(0.310)
R^2 overall	0.132	0.136	0.136	0.154
Wald test	472.772***	607.972***	665.532***	687.710***
Obs.	15810	15810	15810	15810

Robust standard errors in brackets,
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

mentation of decision processes in the human brain follows processes of a more gradual nature (e.g., Shadlen and Kiani, 2013). We have demonstrated the existence of a stable, gradual relation between error rates and an underlying, cardinal "strength of preference," and shown that the latter is best represented by integrated variables of an exclusively economic nature. That is, decisions become more error-prone as the economic distance between the alternatives becomes smaller.

Our research strategy has followed three complementary approaches. First, we have shown that, in decisions in the domain of judgment and decision making where objectively-correct options can be identified, expected value differences are enough to explain error rates. This is important, because such an explanatory variable is independent of any estimation of subjective values and hence constitutes the direct parallel to psychophysical studies which have identified gradual effects as a function of objective differences in weight, length, brightness, etc. The typical candidate explanatory variable derived from purely psychophysical approaches for economic tasks, (payoff-irrelevant) numerical differences (Moyer and Landauer, 1967; Dehaene, Dupoux and Mehler, 1990), does play a role but can be safely considered a second-order variable.

Second, we have shown that, in decisions under risk in the subjective domain, where what is correct depends on individual risk attitudes, strength of preference can be characterized by an integrated variable reflecting differences in expected utility, while expected value differences do a considerably worse job. Again, numerical differences do play a role, but appear to be relatively less important than pure economic distance. Crucially, our approach has followed a strictly out-of-sample procedure where utility functions are estimated on one part of the dataset and the test of gradual effects between utility differences and error rates is conducted using the choices in a different part of the dataset. This is important, because fitting a dataset with, say, a random utility model merely *assumes* that errors follow a smooth distribution; that is, gradual effects are *assumed* and would appear to be present after the fact even if they did not exist at all.

Third, we have shown that the relation between strength of preference, as captured by notions of economic distance, and error rates reflects more than an *ex post* and *as if* model. The same gradual effects are obtained when examining response times, with easier decisions (those where economic distance is large) being made faster than harder ones (those where economic distance is small). Response times are a straightforward, easily-measurable reflection of the actual functioning of human decision processes. Most importantly, they are unrelated to estimation and fitting procedures and hence serve as an independent confirmation of the postulated effects.

Our results provide empirical support and explicit foundation for the literature on stochastic choice, which has been long advocated as a realistic building block for theories of microeconomic decision making (Debreu, 1958; Davidson and Marschak, 1959; Luce, 1959; Machina, 1985). It is in line with modern empirical contributions pointing out the ubiquitousness of stochastic choice and decision inconsistencies (Camerer, 1989; Hey and Orme, 1994; Agranov and Ortoleva, 2017), but goes beyond those by precisely examining the content of elusive concepts as "strength of preference" and "choice difficulty" and isolating them from possibly-artificial phenomena derived, e.g., from the underlying assumptions of models used to estimate noisy utility.

The analysis is broadly in line with the psychophysics and neuroscience literature, where the presence of gradual effects on decision making is regarded as an elementary, firmly-established fact (e.g., Weber's Law), but goes beyond it by showing that economic distance is not as simple as objectively-measurable weight or length. Economic decisions are decisions, and hence it is unsurprising that the same (neural) mechanisms that determine perception-based judgments also play a role in them. However, economic decisions are *complex* decisions, and it is equally unsurprising that simple applications of psychophysics (as, say, taking only numerical magnitudes into account) fall short of the task of accounting for economic errors.

Conceptually, our results agree with earlier studies as Mosteller and Nogee (1951) and with recent contributions as Khaw, Li and Woodford (2018). Both report gradual increases in the proportion of risky choices in lottery experiments as the reward increases. Khaw, Li and Woodford (2018) argue in terms of an imprecise perception of stimuli (Green and Swets, 1966; Ma et al., 2006). Those are payoff-relevant numerical magnitudes and hence aligned with economic distance as we consider it.

It is also important to remark that the sigmoidal relation between economic distance and choice frequencies arises spontaneously from the data, hence providing empirical support for random utility models as typically used in applied microeconomics, which often employ logit or probit error distributions. By taking a step back from fitting approaches, our analysis highlights the presence of a systematic structure of noise terms reflecting the gradual effects of choice difficulty. This observation builds upon earlier arguments by Hey and Orme (1994) and Harless and Camerer (1994), which attempted to shift the focus in microeconomics away from deterministic choice models as alternatives to expected utility theory.[12]

We view the relation we study here as the basic building block underlying errors in economic decision making. A very large literature has studied *heuristics* and *biases* in decision making (Kahneman, Slovic and Tversky, 1982; Kahneman, 2003; Grether, 1980, and many others), which are conceived of as systematic, directional deviations from normatively rational or consistent behavior. An equally large literature has argued that such phenomena can be explained in terms of *dual-process theories* including alternative decision processes of intuitive, impulsive, or heuristic nature (Thaler and Shefrin, 1981; Weber and Johnson, 2009; Alós-Ferrer and Strack, 2014, to mention just a few), and in particular their presence has consequences for both choices and response times. We believe that the gradual effects we describe here and dual-process theories are complementary. Building upon the research presented here, closely-related work shows that intuitive decision processes can be identified even when controlling for the gradual effects described in this work. Specifically, Alós-Ferrer and Farolfi (2019) considers a belief-updating task as in Grether (1980, 1992), which falls into the objective domain, and identifies gradual effects as those seen in Experiment 1, where the explanatory variable is exogenously given. Those effects coexist with additional decision processes reflecting well known probability-judgment biases, namely conservatism and the representativeness heuristic (Kahneman and Tversky, 1972; Gennaioli and Shleifer, 2010). Analogously, Alós-Ferrer, Buckenmaier and Garagnani (2019) considers a lottery-choice task, typical of decision making under risk in the subjective domain, and identifies gradual effects reflecting expected utility differences following an out-of-sample approach as in Experiment 2. Again, those effects can be seen to coexist with an additional decision process reflecting the well-known certainty heuristic (Kahneman and Tversky, 1979; Bordalo, Gennaioli and Shleifer, 2012). In both cases, gradual effects both on choice frequencies and response times can be identified when controlling for the candidate heuristics, and vice versa, demonstrating the complementarity of the approach.

[12] "Perhaps we should now spend some time on thinking about the noise, rather than about even more alternatives to expected utility?" (Hey and Orme, 1994).

Even more important, it needs to be remarked that accounting for the gradual effects arising from economic distance is actually a precondition to be able to properly identify heuristics and alternative decision processes in choice data. The reason is that failing to account for such effects may bias the analysis, leading e.g. to a reverse inference fallacy or to the attribution of patterns in error rates and response times to heuristics when strength of preference might suffice to explain them (Krajbich et al., 2015).

The implications of our results are of broad significance for economic modeling. First, the demonstration of the gradual relation between economic integrated variables and errors provides a foundation for theories of stochastic choice and empirical approaches to preference revelation alike. Second, the fact that these effects are a natural extension of those observed in psychophysics provides a tangible bridge to other disciplines, most notably neuroscience, through which new techniques and ideas can travel. Third, the results pose a significant challenge to traditional, neoclassic modeling. For the latter is based on deterministic and, more importantly, *purely ordinal* preferences. *As if* models can be justified as fitting and prediction exercises, hence compatible with ordinal approaches. Our results on gradual mappings from economic variables to error rates and response times, though, go beyond any *as if* interpretation, and, in our opinion, are best viewed in the context of an inherently cardinal view of preferences.

References

Achtziger, Anja, and Carlos Alós-Ferrer. 2014. "Fast or Rational? A Response-Times Study of Bayesian Updating." *Management Science*, 60(4): 923–938.

Agranov, Marina, and Pietro Ortoleva. 2017. "Stochastic Choice and Preferences for Randomization." *Journal of Political Economy*, 125(1): 40–68.

Alekseev, Aleksandr, Glenn W. Harrison, Morten Lau, and Don Ross. 2019. "Deciphering the Noise: The Welfare Costs of Noisy Behavior." Working Paper.

Alós-Ferrer, Carlos, and Alexander Ritschel. 2018. "The Reinforcement Heuristic in Normal Form Games." *Journal of Economic Behavior and Organization*, 152: 224–234.

Alós-Ferrer, Carlos, and Federica Farolfi. 2019. "Imperfect Bayesians." Working Paper, University of Zurich.

Alós-Ferrer, Carlos, and Fritz Strack. 2014. "From Dual Processes to Multiple Selves: Implications for Economic Behavior." *Journal of Economic Psychology*, 41: 1–11.

Alós-Ferrer, Carlos, Dura-Georg Granić, Johannes Kern, and Alexander K. Wagner. 2016. "Preference Reversals: Time and Again." *Journal of Risk and Uncertainty*, 52(1): 65–97.

Alós-Ferrer, Carlos, Ernst Fehr, and Nick Netzer. 2018. "Time Will Tell: Recovering Preferences when Choices are Noisy." Working Paper, University of Zurich.

Alós-Ferrer, Carlos, Johannes Buckenmaier, and Michele Garagnani. 2019. "A Tale of Two Mechanisms: The Determinants of Errors in Decisions Under Risk." Working Paper, University of Zurich.

Anderson, Simon P., Jacques-François Thisse, and André De Palma. 1992. *Discrete Choice Theory of Product Differentiation.* Cambridge, MA:MIT Press.

Apesteguía, José, and Miguel A. Ballester. 2018. "Monotone Stochastic Choice Models: The Case of Risk and Time Preferences." *Journal of Political Economy*, 126(1): 74–106.

Apesteguía, José, Miguel A. Ballester, and Jay Lu. 2017. "Single-Crossing Random Utility Models." *Econometrica*, 85(2): 661–674.

Ballinger, T. Parker, and Nathaniel T. Wilcox. 1997. "Decisions, Error and Heterogeneity." *Economic Journal*, 107(443): 1090–1105.

Bellemare, Charles, Sabine Kröger, and Arthur van Soest. 2008. "Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities." *Econometrica*, 76(4): 815–839.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2012. "Salience Theory of Choice under Risk." *Quarterly Journal of Economics*, 127(3): 1243–1285.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2013. "Salience and Consumer Choice." *Journal of Political Economy*, 121(5): 803–843.

Camerer, Colin F. 1989. "Does the Basketball Market Believe in the 'Hot Hand'." *American Economic Review*, 79: 1257–1261.

Caplin, Andrew, and Mark Dean. 2015. "Revealed Preference, Rational Inattention, and Costly Information Acquisition." *American Economic Review*, 105(7): 2183–2203.

Chabris, Christopher F., Carrie L. Morris, Dmitry Taubinsky, David Laibson, and Jonathon P. Schuldt. 2009. "The Allocation of Time in Decision-Making." *Journal of the European Economic Association*, 7(2-3): 628–637.

Charness, Gary, and Dan Levin. 2005. "When Optimal Choices Feel Wrong: A Laboratory Study of Bayesian Updating, Complexity, and Affect." *American Economic Review*, 95(4): 1300–1309.

Choi, Syngjoo, Shachar Kariv, Wieland Müller, and Dan Silverman. 2014. "Who Is (More) Rational?" *American Economic Review*, 104(6): 1518–1550.

Conte, Anna, John D. Hey, and Peter G. Moffatt. 2011. "Mixture Models of Choice Under Risk." *Journal of Econometrics*, 162(1): 79–88.

Dashiell, John F. 1937. "Affective Value-Distances as a Determinant of Aesthetic Judgment-Times." *American Journal of Psychology*, 50: 57–67.

Davidson, Donald, and Jacob Marschak. 1959. "Experimental Tests of a Stochastic Decision Theory." In *Measurement: Definitions and Theories.* , ed. West Churchman and Philburn Ratoosh, 233–269. New York:Wiley.

Debreu, Gerard. 1958. "Stochastic Choice and Cardinal Utility." *Econometrica*, 26(3): 440–444.

Deese, James, and Roger A. Kaufman. 1957. "Serial Effects in Recall of Unorganized and Sequentially Organized Verbal Material." *Journal of Experimental Psychology*, 54(3): 180–187.

Dehaene, Stanislas. 1992. "Varieties of Numerical Abilities." *Cognition*, 44(1-2): 1–42.

Dehaene, Stanislas, Emmanuel Dupoux, and Jacques Mehler. 1990. "Is Numerical Comparison Digital? Analogical and Symbolic Effects in Two-Digit Number Comparison." *Journal of Experimental Psychology: Human Perception and Performance*, 16(3): 626–641.

Dehaene, Stanislas, Véronique Izard, Elizabeth Spelke, and Pierre Pica. 2008. "Log or Linear? Distinct Intuitions of the Number Scale in Western and Amazonian Indigene Cultures." *Science*, 320(5880): 1217–1220.

Fechner, T. Gustav. 1860. *Elemente der Psychophysik [Elements of Psychophysics].* Leipzig:Breitkopf & Härtel.

Fischbacher, Urs, Ralph Hertwig, and Adrian Bruhin. 2013. "How to Model Heterogeneity in Costly Punishment: Insights from Responders' Response Times." *Journal of Behavioral Decision Making*, 26(5): 462–476.

Fudenberg, Drew, and Tomasz Strzalecki. 2015. "Dynamic Logit with Choice Aversion." *Econometrica*, 83(2): 651–691.

Fudenberg, Drew, Philipp Strack, and Tomasz Strzalecki. 2018. "Speed, Accuracy, and the Optimal Timing of Choices." *American Economic Review*, 108(12): 3651–3684.

Gabaix, Xavier. 2014. "A Sparsity-Based Model of Bounded Rationality." *Quarterly Journal of Economics*, 129(4): 1661–1710.

Gabaix, Xavier, David Laibson, Guillermo Moloche, and Stephen Weinberg. 2006. "Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model." *American Economic Review*, 96(4): 1043–1068.

Gennaioli, Nicola, and Andrei Shleifer. 2010. "What Comes To Mind." *Quarterly Journal of Economics*, 125(4): 1399–1433.

Goeree, Jacob K., Charles A. Holt, and Thomas R. Palfrey. 2005. "Regular Quantal Response Equilibrium." *Experimental Economics*, 8: 347–367.

Green, David M., and John A. Swets. 1966. *Signal Detection Theory and Psychophysics.* New York, NY:Wiley.

Greiner, Ben. 2015. "Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE." *Journal of the Economic Science Association*, 1: 114–125.

Grether, David M. 1980. "Bayes Rule as a Descriptive Model: The Representativeness Heuristic." *Quarterly Journal of Economics*, 95: 537–557.

Grether, David M. 1992. "Testing Bayes Rule and the Representativeness Heuristic: Some Experimental Evidence." *Journal of Economic Behavior and Organization*, 17: 31–57.

Grether, David M., and Charles R. Plott. 1979. "Theory of Choice and the Preference Reversal Phenomenon." *American Economic Review*, 69(4): 623–638.

Harless, David W., and Colin F. Camerer. 1994. "The Predictive Utility of Generalized Expected Utility Theories." *Econometrica*, 62(6): 1251–1289.

Harrison, Glenn W. 2008. "Maximum Likelihood Estimation of Utility Functions Using Stata." Working Paper.

Harrison, Glenn W., and Jia Min Ng. 2016. "Evaluating the Expected Welfare Gain from Insurance." *Journal of Risk and Insurance*, 83(1): 91–120.

Hey, John D., and Chris Orme. 1994. "Investigating Generalizations of Expected Utility Theory Using Experimental Data." *Econometrica*, 62(6): 1291–1326.

Hicks, John R., and Roy G.D. Allen. 1934. "A Reconsideration of the Theory of Value. Part I." *Economica*, 1(1): 52–76.

Kahneman, Daniel. 2003. "Maps of Bounded Rationality: Psychology for Behavioral Economics." *American Economic Review*, 93(5): 1449–1475.

Kahneman, Daniel, and Amos Tversky. 1972. "Subjective Probability: A Judgment of Representativeness." *Cognitive Psychology*, 3: 430–454.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica*, 47(2): 263–291.

Kahneman, Daniel, Paul Slovic, and Amos Tversky. 1982. *Judgment Under Uncertainty: Heuristics and Biases.* UK:Cambridge University Press.

Khaw, Mel Win, Ziang Li, and Michael Woodford. 2018. "Cognitive Imprecision and Small-Stakes Risk Aversion." Working Paper, Columbia University.

Klein, A. Stanley. 2001. "Measuring, Estimating, and Understanding the Psychometric Function: A Commentary." *Attention, Perception, & Psychophysics*, 63(8): 1421–1455.

Krajbich, Ian, Björn Bartling, Todd Hare, and Ernst Fehr. 2015. "Rethinking Fast and Slow Based on a Critique of Reaction-Time Reverse Inference." *Nature Communications*, 6(7455): 1–9.

Laming, Donald. 1985. "Some Principles of Sensory Analysis." *Psychological Review*, 92(4): 462–485.

Lichtenstein, Sarah, and Paul Slovic. 1971. "Reversals of Preference Between Bids and Choices in Gambling Decisions." *Journal of Experimental Psychology*, 89(1): 46–55.

Loomes, Graham, and Robert Sugden. 1995. "Incorporating a Stochastic Element into Decision Theories." *European Economic Review*, 39(3–4): 641–648.

Loomes, Graham, and Robert Sugden. 1998. "Testing Different Stochastic Specifications of Risky Choice." *Economica*, 65(260): 581–598.

Loomes, Graham, Peter G. Moffatt, and Robert Sugden. 2002. "A Microeconometric Test of Alternative Stochastic Theories of Risky Choice." *Journal of Risk and Uncertainty*, 24(2): 103–130.

Luce, R. Duncan. 1959. *Individual Choice Behavior: A Theoretical Analysis.* New York:Wiley.

Machina, Mark J. 1985. "Stochastic Choice Functions Generated from Deterministic Preferences over Lotteries." *Economic Journal*, 95(379): 575–594.

Manzini, Paola, and Marco Mariotti. 2014. "Stochastic Choice and Consideration Sets." *Econometrica*, 82: 1153–1176.

Marschak, Jacob. 1960. "Binary Choice Constraints on Random Utility Indicators." In *Stanford Symposium on Mathematical Methods in the Social Sciences.* , ed. Kenneth J. Arrow, 312–329. Stanford, CA:Stanford University Press.

Matějka, Filip, and Alisdair McKay. 2015. "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model." *American Economic Review*, 105(1): 272–298.

Ma, Wei Ji, Jeffrey M. Beck, Peter E. Latham, and Alexandre Pouget. 2006. "Bayesian Inference with Probabilistic Population Codes." *Nature Neuroscience*, 9(11): 1432–1438.

McFadden, Daniel L. 2001. "Economic Choices." *American Economic Review*, 91(3): 351–378.

McKelvey, Richard D., and Thomas R. Palfrey. 1995. "Quantal Response Equilibria for Normal Form Games." *Games and Economic Behavior*, 10(1): 6–38.

Moffatt, G. Peter. 2015. *Experimetrics: Econometrics for Experimental Economics.* London:Palgrave Macmillan.

Moffatt, Peter G. 2005. "Stochastic Choice and the Allocation of Cognitive Effort." *Experimental Economics*, 8(4): 369–388.

Moffatt, Peter G., Stefania Sitzia, and Daniel J. Zizzo. 2015. "Heterogeneity in Preferences Towards Complexity." *Journal of Risk and Uncertainty*, 51(2): 147–170.

Mosteller, Frederick, and Philip Nogee. 1951. "An Experimental Measurement of Utility." *Journal of Political Economy*, 59: 371–404.

Moyer, Robert S., and Thomas K. Landauer. 1967. "Time Required for Judgements of Numerical Inequality." *Nature*, 215(5109): 1519–1520.

Murdock Jr., Bennet B. 1962. "The Serial Position Effect of Free Recall." *Journal of Experimental Psychology*, 64(5): 482–488.

Peirce, Jonathan W. 2007. "PsychoPy – Psychophysics Software in Python." *Journal of Neuroscience Methods*, 162(1): 8–13.

Ratcliff, Roger. 1978. "A Theory of Memory Retrieval." *Psychological Review*, 85: 59–108.

Rubinstein, Ariel. 2007. "Instinctive and Cognitive Reasoning: A Study of Response Times." *Economic Journal*, 117(523): 1243–1259.

Rubinstein, Ariel. 2013. "Response Time and Decision Making: An Experimental Study." *Judgment and Decision Making*, 8(5): 540–551.

Shadlen, Michael N., and Daphna Shohamy. 2016. "Decision Making and Sequential Sampling from Memory." *Neuron*, 90: 927–939.

Shadlen, Michael N., and Roozbeh Kiani. 2013. "Decision Making as a Window on Cognition." *Neuron*, 80: 791–806.

Soltani, Alireza, Benedetto De Martino, and Colin Camerer. 2012. "A Range-Normalization Model of Context-Dependent Choice: A New Model and Evidence." *PLoS Computational Biology*, 8(7): 1–15.

Spiliopoulos, Leonidas, and Andreas Ortmann. 2018. "The BCD of Response Time Analysis in Experimental Economics." *Experimental Economics*, 21(2): 383–433.

Spitzer, Bernhard, Leonhard Waschke, and Christopher Summerfield. 2017. "Selective Overweighting of Larger Magnitudes During Noisy Numerical Comparison." *Nature Human Behaviour*, 1(0145): 1–8.

Sürücü, Oktay, Behnud Mir Djawadi, and Sonja Recker. 2019. "The Asymmetric Dominance Effect: Reexamination and Extension in Risky Choice - An Experimental Study." *Journal of Economic Psychology*, 73: 102–122.

Thaler, Richard H., and Hersch M. Shefrin. 1981. "An Economic Theory of Self-Control." *Journal of Political Economy*, 89(2): 392–406.

Thurstone, L. 1927. "A Law of Comparative Judgement." *Psychological Review*, 34: 273–286.

Tversky, Amos. 1969. "Intransitivity of Preferences." *Psychological Review*, 76: 31–48.

Tversky, Amos, Paul Slovic, and Daniel Kahneman. 1990. "The Causes of Preference Reversal." *American Economic Review*, 80(1): 204–217.

Von Gaudecker, Hans-Martin, Arthur Van Soest, and Erik Wengström. 2011. "Heterogeneity in Risky Choice Behavior in a Broad Population." *American Economic Review*, 101(2): 664–694.

Weber, Elke U., and Eric J. Johnson. 2009. "Mindful Judgment and Decision Making." *Annual Review of Psychology*, 60: 53–85.

Wichmann, A. Felix, and N. Jeremy Hill. 2001. "The Psychometric Function: I. Fitting, Sampling, and Goodness of Fit." *Attention, Perception, & Psychophysics*, 63(8): 1293–1313.

Wilcox, Nathaniel T. 1993. "Lottery Choice: Incentives, Complexity, and Decision Time." *Economic Journal*, 103(421): 1397–1417.

Wilcox, Nathaniel T. 1994. "On a Lottery Pricing Anomaly: Time Tells the Tale." *Journal of Risk and Uncertainty*, 8(7): 311–324.

Wilcox, Nathaniel T. 2011. "Stochastically More Risk Averse: A Contextual Theory of Stochastic Discrete Choice Under Risk." *Journal of Econometrics*, 162(1): 89–104.