Improving Risky-Choice Predictions Using Response Times*

Carlos Alós-Ferrer[†1] and Michele Garagnani[2]

[1]Zurich Center for Neuroeconomics, University of Zurich

[2]Department of Finance, University of Melbourne

Abstract

Structural methods to estimate preferences for decisions under risk often perform poorly when predicting new decisions out of sample. We compare standard structural methods to a new nonparametric method which reveals preferences with the help of response times, and does so independently of any utility or (symmetric) noise specification. Using two datasets involving choices over risky prospects, we show that the new method outperforms standard (parametric) utility estimations in terms of predictive accuracy.

JEL Classification: D01 · D81 · D91

Keywords: Predicting Choices · Stochastic Choice · Preference Revelation · Response Times

Forthcoming, *Journal of Political Economy: Microeconomics*

(Accepted November 2023)

*We thank Georg Nöldeke, Rob Ranyard, Ulrich Schmidt, Eyal Winter, and seminar participants at the Paris School of Economics, the University of Basel, and the ESA 2022 and FUR 2022 conferences in Bologna and Ghent for helpful comments. Edited by Eduardo Azevedo (Editor) and Jay Lu (Associate Editor).

†Corresponding author. Zurich Center for Neuroeconomics, Department of Economics, University of Zurich, Blümlisalpstrasse 10, CH-8006 Zurich, Switzerland. E-mail: carlos.alos-ferrer@econ.uzh.ch

1 Introduction

Estimating preferences for decisions under risk is of fundamental importance for economics, finance, management, and other fields. Thus, a plethora of methods for preference elicitation has been developed. Those include multiple-question lists (Holt and Laury, 2002, 2005; Eckel and Grossman, 2002), investment tasks (Gneezy and Potters, 1997), and structural maximum likelihood methods (Andersen et al., 2008; Moffatt, 2015), among others. Unfortunately, the performance of these methods according to a variety of criteria leaves much to be desired. The consistency across different risk-elicitation tasks and methods is generally low, as is the test-retest reliability for any given method (e.g. Crosetto and Filippin, 2016; Csermely and Rabas, 2016; Mata et al., 2018; Holzmeister and Stefan, 2021; Perez, Hollard, and Vranceanu, 2021). That is, preferences estimated according to one method are not a good predictor for the preferences estimated according to another (or the same) method. Most importantly, when one examines choices out of sample (that is, decisions other than the ones used in the estimation procedure), the predictive performance of preferences elicited according to those methods is poor (Dave et al., 2010; Holt and Laury, 2014; Csermely and Rabas, 2016; Beauchamp, Cesarini, and Johannesson, 2017; Charness et al., 2018, 2020). For example, Garagnani (2023) recently studied the predictive performance of several established risk-elicitation tasks and found them not to predict better than expected-value maximization, which of course requires no elicitation at all. This is unsatisfactory and casts doubts on the usefulness of elicited risk preferences. Commenting on this problem, Friedman et al. (2014) went so far as to argue that, due to the limited predictive validity of revealed preference measures, standard preference models as expected utility theory are currently of little use for real-world institutions.

In this work, we show that a newly-introduced method of nonparametric preference revelation has better predictive performance for decisions under risk than currently-used ones. Specifically, we rely on the “Time Will Tell” (TWT) method introduced by Alós-Ferrer, Fehr, and Netzer (2021) for abstract settings, which leverages the information contained in response times to reveal preferences independently of any assumptions on Bernoulli utility functions or distributions of behavioral noise. We then use preferences revealed according to this method to make out-of-sample predictions (that is, for decisions not part of the data used to reveal preferences) and compare the predictive performance to that of standard, structural estimation methods tailored to decisions under risk.

The nonparametric TWT method is based on robust empirical regularities of choices and response times arising in psychology and neuroscience. Specifically, error rates are lower and response times are shorter for easier choice problems than for harder ones. These *psychometric* and *chronometric effects* have been established using discrimination tasks (e.g., which line is longer, which number is larger, etc.; Cattell, 1893, 1902; Moyer and Landauer, 1967; Moyer and Bayer, 1976; Laming, 1985; Dehaene, Dupoux, and Mehler, 1990; Klein, 2001; Wichmann and Hill, 2001), but extend to cases where the correct response is subjective, e.g. favorite colors (Dashiell, 1937). An early contribution by Mosteller and Nogee (1951) suggested that psychometric effects (higher consistency for easier choices) are also present for decisions under risk. Recent contributions by Alós-Ferrer and Garagnani (2021, 2022a,b) have systematically demonstrated the presence of both psychometric and chronometric effects in this domain. In particular, when estimating preferences out of sample, new choices take longer and are more likely to contradict the preferences when the difference in expected utilities is smaller.[1]

Conceptually, the TWT method is based on a generalization of random utility models (RUMS; McFadden, 1974, 2001; Manski, 1977; Anderson, Thisse, and De Palma, 1992), which are also the basis of structural maximum likelihood methods to estimate risk preferences. In these models, choices arise from maximization of a utility function u, but utilities are perturbed by a noise term ε. The TWT method, however, goes beyond approaches using RUMs in three directions. First, while structural methods rely on additional parametric assumptions on the shape of the utility u and the distribution of the noise term ε, the TWT method reveals preferences without making parametric assumptions. That is, the method is agnostic on the shape of the utility and the properties of the noise, and reveals preferences not for a particular shape of utility and distribution of noise, but simultaneously for *all* combinations of utility and noise that fit the data. For example, suppose that decisions in a dataset can be explained by a particular functional form of u plus a specific distribution of noise, but they could also be explained by assuming a different utility function (maybe even expected *value* maximization) and a very different distribution of noise. The TWT method then identifies conditions which allow to reveal preferences not for one utility function, but for both of them simultaneously. This is useful because neither utility nor (especially) noise terms are observable, and parametric assumptions are often imposed for analytical convenience only.

[1]Fudenberg, Strack, and Strzalecki (2018) suggested that the fact that error rates and response times are large when utility differences are small reflects the difficulty in separating the values of the options in the decision maker's brain. This process can be captured through *drift-diffusion models* (e.g., Ratcliff, 1978; Shadlen and Kiani, 2013). Baldassi et al. (2020) provided a characterization of those models in terms of psychometric effects.

Second, the method builds upon a more general version of RUMs than the additive RUMs typically used in the literature. In additive RUMs, utility differences for alternatives x, y are perturbed by noise terms $\varepsilon_x, \varepsilon_y$, and a choice of x over y is assumed to occur if $u(x) + \varepsilon_x > u(y) + \varepsilon_y$, or, equivalently, $u(x) - u(y) > \varepsilon_y - \varepsilon_x$. In the more general version that we use, utility noise is allowed to be pair-specific. That is, choice follows a noisy realization $u(x) - u(y) + \varepsilon_{xy}$, where the term ε_{xy} is *not* assumed to be a difference of alternative-specific noise terms. This encompasses standard (additive) RUMs, but also trembling-hand models (Loomes, Moffatt, and Sugden, 2002), where errors are directly related to the choice pairs, and random parameter models (Loomes and Sugden, 1998; Apesteguía and Ballester, 2018), where a parametric functional form is assumed for the utility term, but the parameter is subject to a random perturbation.[2]

Third, the TWT method incorporates response times, allowing to take into account both psychometric and chronometric effects. Psychometric effects are accounted for because the probability of a choice which goes against the utility difference (hence an error) is larger if $u(x)-u(y)$ is closer to zero, that is, errors are more likely as choice pairs are closer to indifference.[3] The TWT method expands the definition of (generalized) RUMS by adding a *chronometric function* capturing the monotonic relation between (realized) utility differences and response times, i.e. chronometric effects. Preference revelation results, however, do not depend on the exact shape of this function, as long as the chronometric effect holds.

Our focus is on decisions under risk and the comparison between structural estimation methods based on additive RUMs (or random parameter models) and the TWT method. Alós-Ferrer, Fehr, and Netzer (2021) applied the method to a dataset on snack choices and obtained 81% correct predictions out of sample. However, this is outside the risk domain and can obviously not be compared to the structural methods typically used there. That is, the open question at this point is not whether the TWT method performs reasonably well in abstract settings, but rather whether it performs *better* than established, structural methods in economically-relevant domains. To answer this question, we focus on two different datasets on risky choices and perform out-of-sample prediction exercises using a number of standard structural methods and the TWT method, keeping the sets

[2]Random parameter models are particular cases of the generalized RUMs in Alós-Ferrer, Fehr, and Netzer (2021) because, in the former, noise in the parameter can be equivalently written as a pair-specific noise term ε_{xy} as in the latter. However, the resulting (transformed) utility noise distribution will generally be non-symmetric.

[3]This is also true for additive RUMS. Psychometric effects were part of the motivation for the original probit model of Thurstone (1927), which is a particular case of additive RUM.

of decisions to be used for estimation and to be used for prediction constant across methods.

The TWT approach requires datasets where subjects make the same choice multiple times *and* where response times were reliably measured. We obtained two datasets with these characteristics from Davis-Stober, Brown, and Cavagnaro (2015) and Kalenscher et al. (2010). We find that the predictive performance of standard, parametric microeconometric methods is rather modest, but the non-parametric TWT method significantly improves upon them. The predictive performance out of sample is high enough to be useful, and larger than the standard levels reported in the literature. Specifically, 76% and 72% of out-of-sample choices, respectively, are correctly predicted for the two datasets. In comparison, none of the structural methods we consider is able to correctly predict more than 60% and 59% of out-of-sample choices, respectively.

Intuitively, the reason that the TWT method improves upon structural estimations is that the latter rely on specific utility forms (CARA, CRRA, etc.) and distributional assumptions for error terms (logit, probit, etc.). In contrast, the TWT method replaces such assumptions with actual data on the response times of choices, and obtains preference revelation results which hold *for all* utility functions and distributions of errors which fit the data (both choices and response times). Of course, all models are wrong, but the TWT model is “less wrong” since its results hold not just for a particular utility form and distribution of noise, but for a large set of them.

This point is important. A further result in Alós-Ferrer, Fehr, and Netzer (2021) allows to predict choice *frequencies* out of sample at the cost of a stronger assumption, namely that error terms have a Fechnerian structure, as is the case of logit and probit models. Strictly speaking, this is not a parametric assumption, but it brings the TWT method closer to standard structural methods using those error distributions. We again examine the performance of the method and find that, under this additional assumption, its advantage over parametric methods vanishes, although no specific set of parametric assumptions improves upon the TWT estimations in *both* datasets.

The paper is structured as follows. Section 2 briefly reviews the theoretical framework. Section 3 presents and compares the out-of-sample prediction analyses for the two datasets. Section 4 presents the additional results assuming Fechner errors. Section 5 concludes. Further analyses and details are in the (Online) Appendix.

2 Non-Parametric Preference Revelation

We compare the predictive performance of the TWT method and standard parametric (structural) methods. The latter entail the estimation of a utility function with a pre-specified functional form (say, CARA or CRRA) within the context of a random utility model (Anderson, Thisse, and De Palma, 1992; McFadden, 2001) or a random parameter model (Loomes and Sugden, 1998; Apesteguía and Ballester, 2018), with additional, specific distributional assumptions on the shape of the noise (e.g., logit or probit models). Appendix A summarizes the (standard) microeconometric approach we followed for the parametric methods.

In contrast, the TWT approach (Alós-Ferrer, Fehr, and Netzer, 2021) entails the non-parametric estimation of an ordinal preference using response times and choice frequencies, which is possible under the additional (also non-parametric) assumption that the noise term is symmetric. Specifically, for a choice between options x and y, this approach identifies joint conditions on the data (choice frequencies and response times) such that *any* model of decisions based on utilities $u(\cdot)$ and (symmetric) pair-specific noise which rationalizes the data must be such that $u(x) > u(y)$. In particular, this encompasses any random utility model with symmetrically-distributed noise, but also trembling-hand models (e.g. Loomes, Moffatt, and Sugden, 2002) which assume a fixed strict preference plus a pair-specific error.[4] The power of the approach is precisely that there is no need to assume any specific model, utility function, or distribution of errors. Rather, preference revelation obtains independently of the model. However, since no particular utility function is estimated, the method is agnostic about specific parameters as e.g. risk attitudes. It merely reveals an ordinal preference, which is however enough to make predictions out of sample.

2.1 Formal Framework

The formal framework is as follows. Let X be a finite set of options. Denote by $C = \{\{x, y\} \mid x, y \in X, x \neq y\}$ the set of all binary choice problems, and let

[4]Random parameter models assume a parametric utility form and add a noise term to the parameter of the utility function instead of to the utility itself. These models are particular cases of the general random utility models in Alós-Ferrer, Fehr, and Netzer (2021), because the parameter noise can be recast in a pair-specific noise term. However, these models will generally violate the symmetry assumption, because symmetric parameter noise does not translate into symmetric utility noise.

$D \subseteq C$ be the set of choice problems on which we have data.[5] A dataset is modeled as follows.

Definition 1. A *stochastic choice function with response times* (SCF-RT) is a pair of functions (p, f) where

(i) p assigns to each $\{x, y\} \in D$ the frequencies $p(x, y) > 0$ and $p(y, x) > 0$, with the property that $p(x, y) + p(y, x) = 1$, and

(ii) f assigns to each $\{x, y\} \in D$ the strictly positive density functions $f(x, y)$ and $f(y, x)$ on $\mathbb{R}_+$.

That is, $p(x, y)$ is the proportion of the time that a decision maker chose x when offered the binary choice between x and y. The assumption that $p(x, y) > 0$ for all $\{x, y\} \in D$ implies that choice is noisy, that is, every alternative is chosen with at least some small probability. The density $f(x, y)$ describes the distribution of response times conditional on x being chosen, and the corresponding cumulative distribution function is denoted by $F(x, y)$.

Definition 2. A *(symmetric) random utility model with a chronometric function* (symmetric RUM-CF) is a triple $(u, \tilde{v}, r)$ where $u : X \to \mathbb{R}$ is a utility function and $\tilde{v} = (\tilde{v}(x, y))_{(x,y)\in C}$ is a collection of real-valued random variables, with each $\tilde{v}(x, y)$ having a density function $g(x, y)$ on $\mathbb{R}$, fulfilling the following properties:

(RUM.1) $\mathbb{E}[\tilde{v}(x, y)] = u(x) - u(y)$,

(RUM.2) $\tilde{v}(x, y) = -\tilde{v}(y, x)$,

(RUM.3) the support of $\tilde{v}(x, y)$ is connected, and

(RUM.4) noise is symmetric: for all $\delta \geq 0$, $g(x, y)(u(x) - u(y) + \delta) = g(x, y)(u(x) - u(y) - \delta)$.

Further, $r : \mathbb{R}_{++} \to \mathbb{R}_+$ is a continuous function that is strictly decreasing in its argument v whenever $r(v) > 0$, with $\lim_{v\to 0} r(v) = \infty$ and $\lim_{v\to\infty} r(v) = 0$.

The random variables $\tilde{v}(x, y)$ and their densities $g(x, y)$ capture noisy choice. Condition (RUM.1) requires that noise is unbiased (equivalent to assuming mean zero for an additive term $\varepsilon_{xy} = \tilde{v}(x, y) - (u(x) - u(y))$). Condition (RUM.2) reflects that the choice between x and y is the same as the choice between y and

[5] Alós-Ferrer, Fehr, and Netzer (2021) employs the notation (x, y) for a choice, i.e. a choice there is written as an ordered pair, but (x, y) and (y, x) represent the same choice. The difference is purely notational.

x. Condition (RUM.3) is a regularity condition requiring that the support has no gaps. (RUM.4) adds the assumption that noise is symmetric.[6] Last, r represents the chronometric function, which maps realized values of utility differences v into response times $r(|v|)$. Specifically, easier choices (where the value $\tilde{v}(x, y)$ is larger) are faster. Given $\{x, y\} \in C$, the random variable describing the response times predicted by the model conditional on x being chosen is given by

$$\tilde{t}(x, y) = r(|\tilde{v}(x, y)|),$$

conditional on $\tilde{v}(x, y) > 0$.

Note that the conditional distributions of response times $F(x, y)$ and $F(y, x)$ can be very different, that is, we do not impose independence between choice outcome and response time. Empirically, response times do differ depending on which alternative is chosen for a given pair (see, e.g. Luce, 1986, Chapter 6.4.3). In terms of the model, it follows from (RUM.2) that the distributions of response times conditional on either choice in a given pair $\{x, y\}$ would be identical if $u(x) = u(y)$, but will in general differ whenever $u(x) \neq u(y)$. The formal result we rely on guarantees preference revelation for *all* symmetric RUM-CFs which rationalize (explain) the data.

Definition 3. A symmetric RUM-CF $(u, \tilde{v}, r)$ *rationalizes* an SCF-RT (p, f) if

(i) $p(x, y) = \text{Prob}[\tilde{v}(x, y) > 0]$ holds for all $\{x, y\} \in D$, and

(ii) $F(x, y)(t) = \text{Prob}[\tilde{t}(x, y) \leq t \mid \tilde{v}(x, y) > 0]$ holds for all $t > 0$ and all $\{x, y\} \in D$.

In other words, a symmetric RUM-CF (the model) rationalizes an SCF-RT (the data) if it reproduces both the choice frequencies and the conditional response time distributions in the latter. We say that an SCF-RT is *rationalizable* within the class of symmetric RUM-CFs if there exists a symmetric RUM-CF in that class that rationalizes it. The next definition captures preference revelation.

Definition 4. A rationalizable SCF-RT *reveals that x is preferred to y* if all symmetric RUM-CFs that rationalize the SCF-RT satisfy $u(x) \geq u(y)$. It *reveals that*

[6]Theorem 1 of Alós-Ferrer, Fehr, and Netzer (2021) considers preference revelation without assuming symmetric noise. This result, however, only identifies sufficient conditions and hence does not generally reveal complete preferences for a given dataset. The assumption of symmetry is needed to ensure that prediction is always possible. Interestingly, any RUM with symmetric noise, and any drift diffusion model with constant or decreasing boundaries is guaranteed to generate data which fulfill the sufficient conditions mentioned above (Alós-Ferrer, Fehr, and Netzer, 2021, Propositions 4 and 6).

x *is strictly preferred to* y if all symmetric RUM-CFs that rationalize the SCF-RT satisfy $u(x) > u(y)$.

2.2 Preference Revelation Out of Sample

It is immediate that, under symmetric noise, $p(x, y) > p(y, x)$ reveals a strict preference for x over y. To obtain out-of-sample predictions (i.e., predictions for choice pairs $\{a, b\} \notin D$), the idea is to triangulate a preference indirectly through comparisons with a reference option. The intuition is that, if a is preferred to x^* with fast response times, this preference is relatively strong, i.e. $u(a)$ is much larger than $u(x^*)$. If b is preferred to the same x^* with slow response times, this preference is relatively weak, i.e. $u(b)$ is only slightly larger than $u(x^*)$. Even though no conclusion follows from transitivity (as both a and b are preferred to x^*), the cardinality embodied in response times should allow to conclude that $u(a)$ is above $u(b)$, that is, a is preferred to b. Theorem 2 in Alós-Ferrer, Fehr, and Netzer (2021) shows that, however, this intuition is elusive, and the meaning of "fast" and "slow" is subtle. Specifically, for each $\{x, y\} \in D$ with $p(x, y) > p(y, x)$, define $\theta(x, y)$ as the $1/2p(x, y)$-percentile of the response time distribution of x, i.e., $F(x, y)(\theta(x, y)) = \frac{1}{2p(x,y)}$. The quantity $\theta(x, y) > 0$ combines information about choice probabilities and response times, that is, it corresponds to a different percentile for each choice pair. Once one replaces "fast or slow response time" with $\theta(x, y)$, the result fully captures the intuition above. We restate it here spelling out all implicit conditions in Alós-Ferrer, Fehr, and Netzer (2021) for convenience.

Theorem 1 (Alós-Ferrer, Fehr, and Netzer, 2021, Theorem 2)**.** *Let* $\{x, y\} \in C \setminus D$ *and suppose there exists* $x_* \in X$ *such that* $\{x, x_*\}, \{y, x_*\} \in D$. *For symmetric RUM-CFs, a rationalizable SCF-RT reveals a preference between* x *and* y *as follows.*

- *If* $p(x, x_*) \geq \frac{1}{2} \geq p(y, x_*)$, *a preference for* x *over* y *is revealed (strictly if one of the inequalities is strict).*
- *If* $p(x, x_*) \leq \frac{1}{2} \leq p(y, x_*)$, *a preference for* y *over* x *is revealed (strictly if one of the inequalities is strict).*
- *If* $p(x, x_*), p(y, x_*) > \frac{1}{2}$, *a preference for* x *over* y *is revealed if* $\theta(x, x_*) \leq \theta(y, x_*)$ *(strictly if this inequality is strict), and a preference for* y *over* x *is revealed otherwise.*

- *If $p(x, x_*), p(y, x_*) < \frac{1}{2}$, a preference for x over y is revealed if $\theta(x_*, x) \geq \theta(x_*, y)$ (strictly if this inequality is strict), and a preference for y over x is revealed otherwise.*

Hence, by fixing a reference option x_*, one can derive a full preference among all alternatives which have been compared to x_* in a dataset including response times.[7] Note that the first two cases in the theorem follow directly by transitivity and because noise is symmetric. The two remaining cases (which are the ones stated in Alós-Ferrer, Fehr, and Netzer, 2021, Theorem 2) are the interesting ones, as the available data and transitivity have no implication in the absence of our result.

In practice, a full preference over the available options in a dataset with a reference option x_* is obtained as follows. Consider all options x such that $p(x, x_*) > \frac{1}{2}$. These options are ranked (in terms of preference) in the inverse order to that dictated by the quantities $\theta(x, x_*)$, i.e. $x \succ x'$ if and only if $\theta(x, x_*) < \theta(x', x_*)$. Now consider all options x such that $p(x, x_*) < \frac{1}{2}$. These options are ranked (in terms of preference) in the order dictated by the quantities $\theta(x_*, x)$, i.e. $x \succ x'$ if and only if $\theta(x_*, x) > \theta(x_*, x')$. Last, if there is any option such that $p(x, x_*) = \frac{1}{2}$, we obtain that $x \sim x_*$.

To apply this result, we need to estimate the density of the distribution of response times. As in Alós-Ferrer, Fehr, and Netzer (2021), the kernel density estimates were performed in *STATA* using the *akdensity* function, which delivers CDFs as output. We estimate the distribution of log-transformed response times to avoid boundary problems. The estimates use an Epanechnikov kernel with optimally chosen non-adaptive bandwidth. For the case where some choice is made only once out of the total number of repetitions (only a single response time is available) an optimal bandwidth cannot be determined endogenously, so we set it to 0.1, yielding a distribution function close to a step function at the observed response time.

Remark 1. In practice, Theorem 1 allows to derive a prediction under two assumptions. The first is that the underlying noise is symmetric, condition (RUM.4). This is weaker than the structural assumptions imposed in parametric estimation methods, and in particular is always fulfilled whenever errors are assumed to have a Fechnerian structure, as is the case of standard approaches using probit (normally-distributed errors) or logit models (see also Section 4). The second is that, for any choice $\{x, y\}$ which is not available in the dataset, there is another option x_* such

[7] We remark that, even though the method capitalizes on the cardinal information contained in response times, it reveals an *ordinal* ranking of alternatives, i.e. a preference.

that the choices $\{x, x_*\}$ and $\{y, x_*\}$ are available in the dataset. For experimental data, fulfilling this condition is a matter of experimental design. For instance, an efficient way to ensure that this condition is met is to include a reference option x_* such that participants face all choices $\{x, x_*\}$ (a "star" design). This is the case in both datasets we use in Section 3.

2.3 Intuition for the Revelation Result

Theorem 1 relies on the psychometric and chronometric effects mentioned in the introduction. These effects are well-established in psychology and neuroscience, and are receiving increasing attention in economics. The psychometric effect refers to the fact that choices are noisier (and error rates are larger) when alternatives are more similar or, in preference terms, when decision makers are closer to indifference (e.g., Cattell, 1893; Dashiell, 1937; Mosteller and Nogee, 1951; Laming, 1985; Klein, 2001; Wichmann and Hill, 2001; Fudenberg, Strack, and Strzalecki, 2018; Alós-Ferrer and Garagnani, 2022a,b). The chronometric effect is the observation that choices are slower when alternatives are more similar or, again in preference terms, when decision makers are closer to indifference (e.g., Cattell, 1902; Dashiell, 1937; Moyer and Landauer, 1967; Moyer and Bayer, 1976; Dehaene, Dupoux, and Mehler, 1990; Moffatt, 2005; Chabris et al., 2009; Krajbich, Oud, and Fehr, 2014; Krajbich et al., 2015; Fudenberg, Strack, and Strzalecki, 2018; Alós-Ferrer and Garagnani, 2022a,b). The psychometric effect is implicitly incorporated in standard random utility models, because it is more likely that (additive) noise will offset an underlying preference of x over y if the utility difference $u(x) - u(y)$ is small than if it is large. The chronometric effect was first incorporated in random utility models in Alós-Ferrer, Fehr, and Netzer (2021). Both are also standard implications of sequential sampling models from the cognitive sciences as the well-known drift-diffusion model (Ratcliff, 1978; Fudenberg, Strack, and Strzalecki, 2018; Baldassi et al., 2020).

The intuition for Theorem 1 is as follows. Assuming symmetric error distributions, $p(x, y) \geq 1/2$ reveals a preference for x over y. Suppose, however, that one does *not* have data for the choice $\{x, y\}$, but there is a third option x_* such that we do have data for the choices $\{x, x_*\}$ and $\{y, x_*\}$. If choice frequencies reveal that x_* is preferred to one of the options in $\{x, y\}$ but not to the other, by transitivity we obtain preference revelation between x and y. These are the first two (straightforward) cases in Theorem 1.

The interesting cases are when choice frequencies reveal that either x_* is preferred to both x and y or the other way around, and hence transitivity yields no additional information. Suppose that $p(x, x_*), p(y, x_*) > \frac{1}{2}$ and hence both x and y are preferred to x_* (the intuition for the other case is analogous). With purely ordinal preferences, nothing can be said about the relation between x and y. By virtue of the chronometric effect, however, response times reveal the strength of preference, and the latter allow for preference revelation. Suppose, for instance, that response times for the choice $\{x, x_*\}$ are relatively short, while response times for the $\{y, x_*\}$ choice are relatively large. By the chronometric effect, this means that the utility difference $u(x) - u(x_*)$ in any model that rationalizes the data must be relatively large, while the utility difference $u(y) - u(x_*)$ must be relatively small. That is, $u(x) - u(x_*) > u(y) - u(x_*) > 0$ and hence $u(x) > u(y)$, revealing a preference for x over y.

The problem, however, lies in capturing the actual meaning of "relatively fast" and "relatively slow." Theorem 1 provides a revelation result independently of the model of (symmetric) noise and the underlying utility function u. This level of generality implies that fast and slow cannot be captured with simple summary statistics as means or medians. The theorem identifies a statistic which combines information from choice frequencies and response times. Specifically, the choice $\{x, x_*\}$ is relatively faster than the choice $\{y, x_*\}$ if $\theta(x, x_*) \leq \theta(y, x_*)$, where $\theta(z, x_*)$ is the $1/2p(z, x_*)$-percentile of the response time distribution for the choice $\{z, x_*\}$ conditional on z being chosen. To see how $\theta(z, x_*)$ combines information from response times and choice frequencies, suppose the latter carry relatively little information, i.e. $p(z, x_*)$ is close to (but above) $1/2$. Then $\theta(z, x_*)$ is a large percentile, becoming the maximum value of the distribution in the limit (if there is one). On the contrary, suppose the choice frequency carries a large amount of information, i.e. $p(z, x_*)$ is close to one. Then $\theta(z, x_*)$ approaches the median. That is, $\theta(z, x_*)$ becomes a larger percentile of the (conditional) response time distribution as choice frequencies carry less and less information. The key is that the comparison between $\theta(x, x_*)$ and $\theta(y, x_*)$ involves different percentiles of the respective response time distributions reflecting the fact that the respective choice frequencies carry different amounts of information for the two choice pairs.

3 Predicting Choices Out of Sample

3.1 Description of the Datasets

In this section we conduct out-of-sample prediction analyses using both standard structural models and Theorem 1. We rely on two existing datasets, from Davis-Stober, Brown, and Cavagnaro (2015) (DSBC) and Kalenscher et al. (2010)

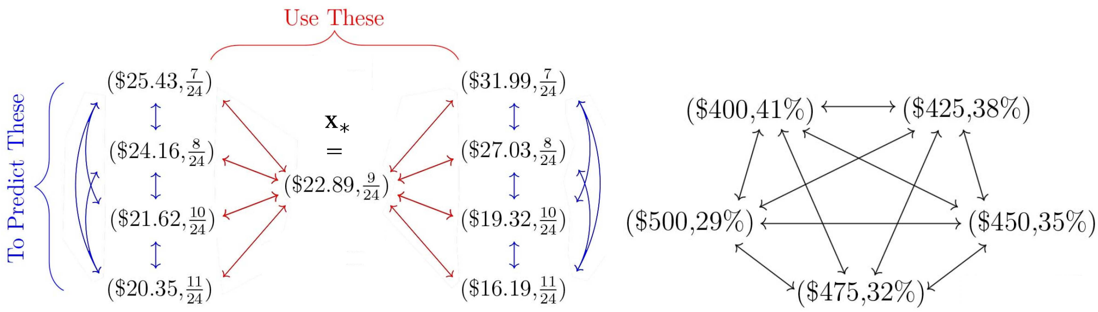

Figure 1: The datasets of DSBC (left) KTHDP (right). Assuming symmetric noise in a dataset with a reference option (x_*) we can use all comparisons with this reference (in red for DSBC) to predict out-of-sample all other choices (in blue). For KTHDP, this can be done for any of the lotteries.

(KTHDP). Those are ideal for our purposes because they include response times and every participant repeated every choice a reasonable number of times.

In the dataset of DSBC, $N = 60$ subjects made binary choices among different lotteries in a 2×2 within-subject design. Specifically, the experiment varied the display format of the lotteries (pies vs. bars) and whether participants faced a time constraint or not (4 seconds vs. no time limit). The choice pairs were drawn from two sets of five lotteries each, with one lottery common to the two sets. All possible combinations of the lotteries within each set were implemented, giving rise to 20 distinct choice pairs. Each of these pairs was repeated 12 times in each of the 4 possible conditions, for a total of $12 \times 4 \times 20 = 960$ choices per participant. Each participant took part in two sessions, each with two (randomly allocated) combinations of time pressure and display format manipulations. One decision from each condition was randomly selected and paid.

In the dataset of KTHDP, $N = 30$ subjects made all possible binary choices among five different lotteries. Each of the 10 resulting choice pairs was repeated 20 times, for a total of 200 trials per participant. Participants needed to decide within 4 seconds (otherwise the trial was missed). Each participant took part in a single session while being scanned in an fMRI scanner. One randomly-selected decision was paid, with dummy dollars converted into Euro at a 100:1 rate.

The structure of the dataset of DSBC (Figure 1, left) allows for a direct application of Theorem 1. All lotteries were repeatedly compared to a specific one (denoted x_* in the figure, where these comparisons are highlighted in red). Theorem 1 above then allows to estimate a full preference relation using just those decisions. The dataset also includes choices within the left-hand and right-hand subsets (highlighted in blue in Figure 1). The estimated preferences can then be used to predict the latter choices. That is, we can use the choice frequencies and response times of the first type of choices (in red) to predict all other comparisons (in blue) out of sample.

For structural estimation, we follow standard econometric approaches. As frequently done in the literature, we assume an additive RUM with either a CRRA utility function or a CARA function. We then repeat the analysis assuming a random parameter model (with either CRRA or CARA functions) instead. For each of the four resulting structural methods, we estimated risk attitudes, separately for each individual, based only on the choices which involved x_* (see Appendix A for details on the estimation procedures). We then used this individual estimate to predict all other choices not involving the lottery x_*. That is, all procedures estimate preferences using the same first set of choices (in red in Figure 1) and predict choices from the same second set (in blue).

In the dataset of KTHDP, all binary choices among five different lotteries were made (Figure 1, right). To implement a comparable out-of-sample approach, we replicated the structure of the analysis described above (both for the structural methods and for Theorem 1) five times, with each analysis adopting one of the five distinct lotteries in KTHDP as reference lottery x_*. For example, we applied Theorem 1 and estimated utilities with a standard microeconometric approach using only the four binary choices involving option [\$500; 29%] and then predicted the remaining six comparisons not involving this option. We did this for each possible lottery, and report the average predictive performance across the five different analyses.

3.2 Results

Figure 2 illustrates the results for both datasets. For DSBC, the predictive performance of TWT, measured as the average out-of-sample proportion of correctly predicted choices, is 76.14% (median 77.13%, min 49.62%, max 100%). This is a reasonably-high performance.[8] More importantly, the out-of-sample predictive performance of TWT is significantly higher than that of standard econometric techniques (Figure 2, left). Crucially, this observation holds independently of the particular utility function assumed (CRRA vs. CARA) and of assumptions on the shape of the noise (RUM vs. RPM). For DSBC (Figure 2, left) the average out-of-sample proportion of correctly predicted choices according to RUM (CRRA) is 57.78% (median 56.25%, min 31.25%, max 87.50%) which is signifi-

[8] Alós-Ferrer, Fehr, and Netzer (2021) reports an out-of-sample predictive performance of 80.7% in a food choice study where the options were simple food snacks; Garagnani (2023) finds that the out-of-sample predictive performance of different risk elicitation tasks is below 68%.

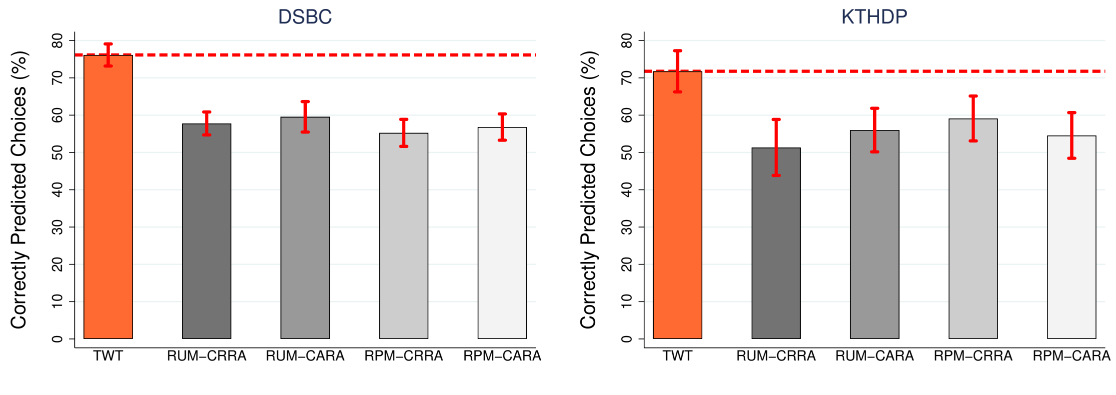

Figure 2: Proportion of out-of-sample correctly predicted choices for RUM/RPM and TWT for Davis-Stober, Brown, and Cavagnaro (2015) (on the left) and Kalenscher et al. (2010) (on the right) across different utility functions (CRRA vs. CARA). 95% confidence intervals are represented in red.

cantly worse than what TWT achieves (WRS, $N = 60$, $z = -6.040$, $p < 0.0001$). The superiority of TWT holds independently of the particular utility function assumed (CRRA vs. CARA) and the assumptions on the shape of the noise (RUM vs. RPM), even accounting for multiple-test corrections (RUM-CARA 59.55%, $z = 7.697$, $p < 0.0001$; RPM-CRRA 55.24%, $z = 8.713$, $p < 0.0001$; RPM-CARA 56.81%, $z = 8.740$, $p < 0.0001$).[9]

The overall picture is very similar for KTHDP's dataset (Figure 2, right). TWT again achieves a reasonable predictive performance (mean 71.76%, median 77.83%, min 32.67%, max 100%) and outperforms standard econometric approaches. The average out-of-sample proportion of correctly predicted choices according to a RUM with CRRA utility and normally-distributed errors is 51.33% (median 50.00%, min 0.00%, max 100.00%) which is significantly smaller than that of TWT (WRS, $N = 26$, $z = -3.881$, $p < 0.001$). A similar result is obtained for the other comparisons (RUM-CARA 56.00%, $z = -3.188$ $p = 0.0008$; RPM-CRRA 59.11%, $z = 3.087$, $p = 0.0013$; RPM-CARA 54.56%, $z = 3.506$, $p = 0.0002$), and accounting for multiple-test corrections.

[9]In Appendix B we show that these results are robust across the different conditions and manipulations in DSBC (time pressure and lottery format).

4 Predictive Performance Assuming Fechner Errors

Standard microeconometric approaches often involve either probit models, i.e. normally-distributed errors, or logit models. Both cases correspond to Fechnerian errors (e.g., Moffatt, 2015), i.e. a fixed shape of the (symmetric) distribution of noise around the utility difference of each pair. Formally, noise in a RUM-CF is *Fechnerian* if, for each $\{x, y\} \in C$ and all $v \in \mathbb{R}$, $g(x, y)(v) = g(v - u(x) + u(y))$, where g is a common density with full support such that $g(\delta) = g(-\delta) > 0$ for all $\delta \geq 0$.

Under the additional assumption of Fechnerian errors (as in any logit or probit model), an additional result of Alós-Ferrer, Fehr, and Netzer (2021) (Theorem 3 there) provides a method to predict the proportion of choices and not just the binary relation, without assuming a specific functional form for utilities or a specific functional shape of the noise term, but assuming that noise is Fechnerian. In this sense, in this section we are still agnostic regarding the specific utility function, but we impose a stronger assumption on the noise. Strictly speaking, this assumption is still nonparametric, but it is closer to the standard parametric assumptions of structural models.

The result is as follows. Say that a RUM-CF is rationalizable in the class of Fechnerian RUM-CFs if it is rationalizable as in Definition 3 according to some Fechnerian RUM-CF (instead of just a symmetric RUM-CF). If an SCF-RT is rationalizable in this sense, we say that it *predicts choice probability* $\bar{p}(x, y)$ for a non-observed choice $\{x, y\} \in C \setminus D$ if all Fechnerian RUM-CFs that rationalize it satisfy $\text{Prob}[\tilde{v}(x, y) > 0] = \bar{p}(x, y)$. That is, the predicted probabilities are independent of the choice of (Fechnerian) RUM-CF, and hence the Fechnerian assumption allows to predict choice probabilities out of sample. The following result pins down the exact predictions.

Theorem 2 (Alós-Ferrer, Fehr, and Netzer, 2021, Theorem 3)**.** *Let* $\{x, y\} \in C \setminus D$ *and* $x_* \in X$ *with* $\{x, x_*\}, \{y, x_*\} \in D$*. Within the class of Fechnerian RUM-CFs, a rationalizable SCF-RT predicts the choice probability*

$$\bar{p}(x, y) = \begin{cases} p(x, z)F(x, z)(\theta(y, z)) & \textit{if } p(y, z) > 1/2, \\ p(x, z) & \textit{if } p(y, z) = 1/2, \\ 1 - p(z, x)F(z, x)(\theta(z, y)) & \textit{if } p(y, z) < 1/2. \end{cases}$$

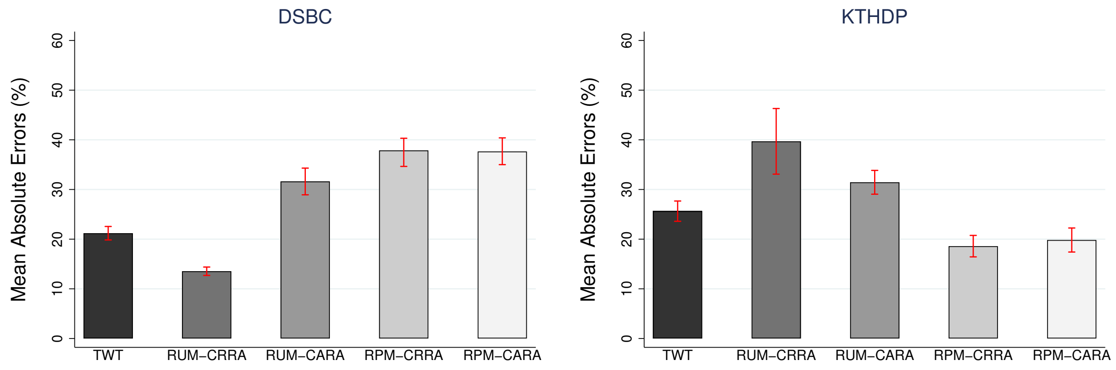

Figure 3: Mean absolute errors for RUM/RPM and TWT for Davis-Stober, Brown, and Cavagnaro (2015) (on the left) and Kalenscher et al. (2010) (on the right) across different utility functions (CRRA vs. CARA). 95% confidence intervals are represented in red.

To apply this result, we follow the same approach as for the previous analysis. That is, for DSBC we predict choice frequencies in decisions not involving x_* (recall Figure 1) after revealing preferences from the decisions involving x_*. For KTHDP, again we average the five possible out-of-sample exercises (taking each distinct lottery in the dataset as the reference).

The standard structural estimations we carried out also allow to predict choice frequencies (instead of deterministic binary choices), making the prediction comparable with Theorem 2. For this purpose, we use the previous estimates to compute the predicted choice frequencies in the corresponding RUM or RPM models. That is, we use the estimated risk attitude and noise variance (for the RUM case) or the estimated mean and variance of the individual risk attitudes (for the RPM case) to predict choice frequencies. See Appendix A for further details.

To measure the accuracy of our predictions, we use the mean absolute error as in Alós-Ferrer, Fehr, and Netzer (2021) and Clithero (2018). This measure calculates the individual average distance between predicted and observed choice frequencies. However, all results are qualitatively unchanged when we use squared errors as a metric of comparison.

The results are shown in Figure 3 (recall that a good performance corresponds to a small mean absolute error). For DSBC, the average mean absolute error across individuals for the TWT method is 0.2120.[10] This outperforms the results when using a RUM estimation with CARA utilities (0.3316; WRS, $N = 60$, $z = -5.926$, $p < 0.0001$) or an RPM approach with either CRRA (0.3787; WRS, $N = 60$, $z =$

[10] Alós-Ferrer, Fehr, and Netzer (2021) obtain 0.237 with the food choice data of Clithero (2018). The latter obtains 0.209 using a parametric drift-diffusion model approach.

-6.618, $p < 0.0001$) or CARA utilities (0.3765; WRS, $N = 60$, $z = -6.530$, $p < 0.0001$). However, the performance of a RUM estimation using CRRA functions is better than that of TWT for this dataset (0.1354; WRS, $N = 60$, $z = 6.140$, $p < 0.0001$).

A qualitatively similar result is obtained for KTHDP's dataset (Figure 3, right). The TWT approach achieves an average mean absolute error of 0.2564, which outperforms RUM estimations with either CRRA (0.3970; WRS, $N = 29$, $z = 2.240$, $p = 0.0240$) or CARA utilities (0.3145; WRS, $N = 24$, $z = 2.286$, $p = 0.0211$). However, RPM estimations outperform TWT for this dataset, both with CRRA (0.1859; WRS, $N = 30$, $z = -2.910$, $p = 0.0028$) and with CARA utilities (0.1982; WRS, $N = 30$, $z = -2.088$, $p = 0.0364$), although these results do not reach significance if adjusting for multiple testing.

The mixed results for this latter analysis might simply reflect the dangers of additional, possibly-unwarranted assumptions used in estimation procedures. Contrary to the previous section, we are now assuming Fechner errors in the TWT approach (although not a specific functional shape), an assumption that might be less warranted than simply symmetric errors as in the first analysis. Fechner errors reduce the conceptual distance between the TWT method and the RUM approach (with normally-distributed errors) or RPM analyses (with normally-distributed risk attitudes). While without this assumption the TWT method significantly outperformed the other, parametric approaches, adopting this assumption leads to mixed results, which are also inconsistent across datasets.

5 Discussion

We show that, for decisions under risk (lottery choice), non-parametric methods using response times have a better predictive performance (of around 75% of correctly-predicted choices out of sample in our datasets) compared to standard parametric, structural estimations. The power of the technique, and the reason why it improves upon structural methods, is that the use of response times allows to *reveal* (ordinal) preferences in a nonparametric fashion, without unwarranted assumptions on the functional form of utilities or the shape of the distribution of noise. Of course, we do not claim that this method should substitute established techniques, as it has its own shortcomings in terms of widespread applicability (e.g., it requires datasets with repeated choices and reference options), but it has the potential to be a powerful tool to validate and improve preference estimation.

The second part of our analysis delivers an important *caveat*. As soon as we introduce a specific, but commonly used, assumption over the distribution of noise, the method's performance for predicting choice frequencies (rather than just choices) is not systematically different from that of standard parametric approaches. This suggests that assumptions over the distribution of the noise, which are often made out of analytical convenience, might sometimes carry even more weight than those on the specific utility functions.[11] Taken together, our empirical results highlight the danger of uncritically using untested assumptions on the distribution of noise.

Data Availability

The data used in the paper is from Davis-Stober, Brown, and Cavagnaro (2015) and Kalenscher et al. (2010). We are grateful to those researchers for sharing the data. Anybody using this data must acknowledge the sources and reference Davis-Stober, Brown, and Cavagnaro (2015) and Kalenscher et al. (2010). Details concerning the experiments can be found in the respective papers. The data and the code replicating the results, tables, and figures in this article can be found in the Harvard Dataverse (Alós-Ferrer and Garagnani, 2023).

References

Alós-Ferrer, Carlos, Ernst Fehr, and Nick Netzer. 2021. "Time Will Tell: Recovering Preferences when Choices are Noisy." *Journal of Political Economy* 129 (6):1828–1877.

Alós-Ferrer, Carlos and Michele Garagnani. 2021. "Choice Consistency and Strength of Preference." *Economics Letters* 198:109672.

———. 2022a. "Strength of Preference and Decisions Under Risk." *Journal of Risk and Uncertainty* 64 (3):309–329.

———. 2022b. "The Gradual Nature of Economic Errors." *Journal of Economic Behavior and Organization* 200:55–66.

Alos-Ferrer, Carlos and Michele Garagnani. 2023. "Replication Data for Improving Risky-Choice Predictions Using Response Times." URL `https://doi.org/10.7910/DVN/XJMHEI`.

Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutström. 2008. "Eliciting Risk and Time Preferences." *Econometrica* 76 (3):583–618.

[11] This resonates with the warning of Hey and Orme (1994): "Perhaps we should now spend some time on thinking about the noise, rather than about even more alternatives to EU?"

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.

Baldassi, Carlo, Simone Cerreia-Vioglio, Fabio Maccheroni, and Massimo Marinacci. 2020. "A Behavioral Characterization of the Drift Diffusion Model and its Multi-Alternative Extension to Choice under Time Pressure." *Management Science* 66 (11):5075–5093.

Beauchamp, Jonathan P., David Cesarini, and Magnus Johannesson. 2017. "The Psychometric and Empirical Properties of Measures of Risk Preferences." *Journal of Risk and Uncertainty* 54 (3):203–237.

Cattell, James McKeen. 1893. "On Errors of Observation." *The American Journal of Psychology* 5 (3):285–293.

———. 1902. "The Time of Perception as a Measure of Differences in Intensity." *Philosophische Studien* 19:63–68.

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, Catherine Eckel, Uri Gneezy, and Agne Kajackaite. 2018. "Complexity in Risk Elicitation May Affect the Conclusions: A Demonstration Using Gender Differences." *Journal of Risk and Uncertainty* 56 (1):1–17.

Charness, Gary, Thomas Garcia, Theo Offerman, and Marie Claire Villeval. 2020. "Do Measures of Risk Attitude in the Laboratory Predict Behavior Under Risk in and Outside of the Laboratory?" *Journal of Risk and Uncertainty* 60 (2):99–123.

Clithero, John A. 2018. "Improving Out-of-Sample Predictions Using Response Times and a Model of the Decision Process." *Journal of Economic Behavior and Organization* 148:344–375.

Crosetto, Paolo and Antonio Filippin. 2016. "A Theoretical and Experimental Appraisal of Four Risk Elicitation Methods." *Experimental Economics* 19 (3):613–641.

Csermely, Tamás and Alexander Rabas. 2016. "How to Reveal People's Preferences: Comparing Time Consistency and Predictive Power of Multiple Price List Risk Elicitation Methods." *Journal of Risk and Uncertainty* 53 (2–3):107–136.

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

Dave, Chetan, Catherine C. Eckel, Cathleen A. Johnson, and Christian Rojas. 2010. “Eliciting Risk Preferences: When Is Simple Better?” *Journal of Risk and Uncertainty* 41 (3):219–243.

Davis-Stober, Clintin P., Nicholas Brown, and Daniel R. Cavagnaro. 2015. “Individual Differences in the Algebraic Structure of Preferences.” *Journal of Mathematical Psychology* 66:70–82.

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.

Eckel, Catherine C. and Philip J. Grossman. 2002. “Sex Differences and Statistical Stereotyping in Attitudes Toward Financial Risk.” *Evolution and Human Behavior* 23 (4):281–295.

Friedman, Daniel, R. Mark Isaac, Duncan James, and Shyam Sunder. 2014. *Risky Curves: On the Empirical Failure of Expected Utility.* New York: Routledge.

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

Garagnani, Michele. 2023. “The Predictive Power of Risk Elicitation Tasks.” *Journal of Risk and Uncertainty* forthcoming.

Gneezy, Uri and Jan Potters. 1997. “An Experiment on Risk Taking and Evaluation Periods.” *Quarterly Journal of Economics* 112 (2):631–645.

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

Holt, Charles A. and Susan K. Laury. 2002. “Risk Aversion and Incentive Effects.” *American Economic Review* 92 (5):1644–1655.

———. 2005. “Risk Aversion and Incentive Effects: New Data Without Order Effects.” *American Economic Review* 95 (3):902–904.

———. 2014. “Assessment and Estimation of Risk Preferences.” In *Handbook of the Economics of Risk and Uncertainty*, edited by Mark J. Machina and Kip W. Viscusi. Oxford, UK: North-Holland, 135–201.

Holzmeister, Felix and Matthias Stefan. 2021. “The Risk Elicitation Puzzle Revisited: Across-Methods (In)consistency?” *Experimental Economics* 24 (2):593–616.

Kalenscher, Tobias, Philippe N. Tobler, Willem Huijbers, Sander M. Daselaar, and Cyriel Pennartz. 2010. “Neural Signatures of Intransitive Preferences.” *Frontiers in Human Neuroscience* 4 (49):1–14.

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.

Krajbich, Ian, Bastiaan Oud, and Ernst Fehr. 2014. "Benefits of Neuroeconomic Modeling: New Policy Interventions and Predictors of Preference." *American Economic Review (Papers and Proceedings)* 104 (5):501–506.

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

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.

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

Luce, R. Duncan. 1986. *Response Times: Their Role in Inferring Elementary Mental Organization*. New York: Oxford University Press.

Manski, Charles F. 1977. "The Structure of Random Utility Models." *Theory and Decision* 8 (3):229–254.

Mata, Rui, Renato Frey, David Richter, Jürgen Schupp, and Ralph Hertwig. 2018. "Risk Preference: A View from Psychology." *Journal of Economic Perspectives* 32 (2):155–172.

McFadden, Daniel L. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*, edited by P. Zarembka. New York: Academic Press, 105–142.

———. 2001. "Economic Choices." *American Economic Review* 91 (3):351–378.

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

———. 2015. *Experimetrics: Econometrics for Experimental Economics*. London: Palgrave Macmillan.

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

Moyer, Robert S. and Richard H. Bayer. 1976. "Mental Comparison and the Symbolic Distance Effect." *Cognitive Psychology* 8 (2):228–246.

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

Perez, Fabien, Guillaume Hollard, and Radu Vranceanu. 2021. "How Serious is the Measurement-Error Problem in Risk-Aversion Tasks?" *Journal of Risk and Uncertainty* 63 (3):319–342.

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

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

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

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.