

Time Will Tell: Recovering Preferences When Choices Are Noisy

Carlos Alós-Ferrer

University of Zurich

Ernst Fehr

University of Zurich

Nick Netzer

University of Zurich

When choice is stochastic, revealed preference analysis often relies on random utility models. However, it is impossible to infer preferences without assumptions on the distribution of utility noise. We show that this difficulty can be overcome by using response time data. A simple condition on response time distributions ensures that choices reveal preferences without distributional assumptions. Standard models from economics and psychology generate data fulfilling this condition. Sharper results are obtained under symmetric or Fechnerian noise, where response times allow uncovering preferences or predicting choice probabilities out of sample. Application of our tools is simple and generates remarkable prediction accuracy.

We thank the editor Emir Kamenica and five anonymous referees for very helpful comments. We are particularly grateful to John Clithero for sharing his data. We also thank Drew Fudenberg, Ian Krajbich, Jay Lu, Paola Manzini, Yusufcan Masatlioglu, Paulo Natenzon, Pietro Ortoleva, Antonio Penta, Antonio Rangel, Jakub Steiner, Tomasz Strzalecki, Ryan Webb, and Michael Woodford as well as seminar audiences at Bocconi University, the University of Bonn, Princeton University, Washington University in St. Louis, the Workshop on Neuroeconomics and the Evolution of Economic Behavior (Vancouver 2018), the Workshop on Stochastic Choice (Barcelona 2018), the Sloan-Nomis Workshop on the Cognitive Foundations of Economic Behavior (Vitznau 2018), the CESifo Area Conference on Behavioral

Electronically published April 28, 2021

[*Journal of Political Economy*, 2021, vol. 129, no. 6]

© 2021 by The University of Chicago. All rights reserved. 0022-3808/2021/12906-00XX\$10.00

I. Introduction

Revealed preference arguments lie at the foundation of economics (e.g., Samuelson 1938; Houthakker 1950; Arrow 1959). Preferences revealed by choice are used in positive economics to predict behavior in novel situations and in normative economics to evaluate the desirability of economic policies.

The traditional revealed preference approach assumes that typical choices are deterministic. Instances of stochastic choice are interpreted as a consequence of indifference, and information about choice frequencies is ignored. These assumptions are challenged by real-world choice behavior, as argued already in the classical work of Fechner (1860) and Luce (1959). Extensive evidence shows that individuals very often make different choices when confronted with the same set of options repeatedly, and their choice frequencies display systematic patterns (among many others, see Tversky 1969; Camerer 1989; Hey and Orme 1994; Agranov and Ortoleva 2017). In view of this evidence, the traditional approach has been modified to explicitly account for stochastic choice and frequency information. The dominant paradigm in applied microeconomics today is to add a random component to cardinal utility (e.g., Thurstone 1927; Marschak 1960; McFadden 1974, 2001). Random utility has several different interpretations, including noise in an individual's perception of the options, temporary fluctuations of tastes, and unobserved heterogeneity in a population of agents. With assumptions on the distribution of the random utility component, it becomes possible to deduce an underlying deterministic utility function from choice behavior.¹

A problem with the random utility approach is that the distributional assumptions may actually drive the results. It is a well-known (but rarely stated) fact that within this approach, nothing can be learned about preferences without making distributional assumptions. The flip side of this result is that anything can be learned by making the suitable assumptions on the structure of noise. Unfortunately, these assumptions cannot be verified, because utility is a latent variable that is not directly observed. Hence, the

Economics (Munich 2018), and the 6th Annual Conference of the Society for Economic Measurement (Frankfurt 2019). C.A.-F. gratefully acknowledges financial support from the Swiss National Science Foundation under project 100014_179009. Data are provided as supplementary material online.

¹ There are alternative approaches to revealed preference when choice is stochastic, which we do not consider here. One is to work with a distribution over ordinal preferences (Block and Marschak 1960; Falmagne 1978; Barberá and Pattanaik 1986; Gul and Pesendorfer 2006; Gul, Natenzon, and Pesendorfer 2014; Apesteguía, Ballester, and Lu 2017; Apesteguía and Ballester 2018). Another one is to model stochastic choice as the result of stochastic consideration sets (e.g., Manzini and Mariotti 2014). See sec. VI for a more detailed literature review.

dependence on possibly unwarranted assumptions is not just an abstract theoretical problem but is known to plague empirical research (see Hey and Orme 1994). For instance, Buschena and Zilberman (2000) showed that for the same data set, assuming homoskedasticity supports non-expected utility models, while expected utility models cannot be improved upon when heteroskedasticity is allowed.

In this paper, we show that the problem can be overcome by using data on response times. This is because the distribution of response times, which is in principle observable, contains information about the unobservable distribution of utility.² We derive, first, a simple and intuitive condition on the distribution of response times that ensures that preferences can be identified from choice data without any assumptions on the structure of noise. Second, we show that under symmetric noise, response times enable the identification of preferences between alternatives for which no previous choice data exist. This would not be possible without response time data unless one is willing to impose stronger (untestable) assumptions than symmetry on the utility noise. Third, we show that if one is willing to assume that utility noise is Fechnerian—an underlying assumption of probit and logit models—response time data enable the calculation of precise choice probabilities for alternatives for which no choice data exist. Again, this would not be possible without response time data under only the Fechnerian assumption.

Our approach is made possible by the fact that, despite being stochastic, choice behavior obeys certain well-known regularities. One of those regularities, often referred to as the *psychometric function*, is the fact that easier choice problems are more likely to elicit correct responses than harder problems. This can be traced back to perceptual discrimination experiments in psychophysics, where an objectively correct response exists (e.g., choosing the brightest or loudest stimulus). It is perhaps one of the most robust facts in all of psychology that the percentage of correct choices increases with the difference in stimuli (Cattell 1893; Laming 1985; Klein 2001; Wichmann and Hill 2001). Conversely, choice becomes noisier when stimuli are more similar, and hence the problem is harder. This finding extends to cases where the correct response is subjective (e.g., favorite colors) and is uncovered by the researcher through ratings (Dashiell 1937). In economics, the classical work of Mosteller and Nogee (1951) showed that the phenomenon also occurs in decisions under risk. In their data, the alternative with the larger estimated utility was not always chosen, but the percentage of choices in favor of the high-utility option

² The time it takes to make a decision (response time) can always be observed in laboratory experiments, but even outside the laboratory response times are in principle an observable outcome of a choice process that can be collected by researchers, firms, or other interested parties. In contrast, the distribution of utility noise is intrinsically unobservable.

increased in the utility difference between the options. This corresponds exactly to the psychometric function, with the difficulty of the (binary) choice problem being measured by the subjective utility difference between the available options and easier choices being those with a larger absolute utility difference. In fact, this psychometric relationship is an integral part of standard random utility models (RUMs), which assume that choice probabilities are monotone in utility differences.

Our approach in this paper rests on integrating a second well-known regularity—often referred to as the *chronometric function*—into the standard random utility framework. The chronometric function describes the fact that easier choice problems take less time to respond than harder problems. As in the case of the psychometric function, there is overwhelming evidence from psychophysics showing this regularity in an astonishing variety of domains, going back to as early as Cattell (1902). For instance, Moyer and Landauer (1967) demonstrated that chronometric effects exist even for the simple question of which of two single-digit numbers is larger (see also Moyer and Bayer 1976; Dehaene, Dupoux, and Mehler 1990). The finding extends to choice based on subjective preferences, as in the work on favorite colors by Dashiell (1937). A growing body of evidence shows that the chronometric phenomenon also applies to economic decisions; that is, response times are decreasing in utility differences. For instance, Chabris et al. (2009) measured response times in binary intertemporal choices and found strong evidence that decisions are faster when utility differences are larger. Krajbich et al. (2015) found the same relation for sharing decisions in a dictator game. Moffatt (2005) and Alós-Ferrer and Garagnani (2018) established the phenomenon for decisions under risk. Krajbich et al. (2012) showed that consumer purchases display the same pattern, and the phenomenon has been repeatedly illustrated in the domain of food choice (Krajbich, Armel, and Rangel 2010; Krajbich and Rangel 2011; Fisher 2017; Clithero 2018). These different papers use very different methods to elicit utility differences—such as stated liking ratings (Krajbich and Rangel 2011), monetary differences (Alós-Ferrer and Garagnani 2018), and estimation based on a logit or probit specification (Moffatt 2005; Chabris et al. 2009; Krajbich et al. 2015; Alós-Ferrer and Garagnani 2018)—but they all arrive at the chronometric relationship between utility differences and response times.

To provide an intuition for our results, consider the choice between two options x and y , where x is chosen with probability p and y with probability $1 - p$. For the sake of clarity, let us adopt the interpretation that these probabilities describe the choices of a single individual across many repetitions of the problem. We will first show that observing $p > 1/2$ is not sufficient to conclude that the individual prefers x to y —that is, that the underlying deterministic utility of x is larger than that of y —if no assumptions about the shape of the distribution of utility noise are made. It is

possible to rationalize the data by a RUM that has a deterministic utility function with $u(x) < u(y)$ and asymmetric noise with zero mean. The asymmetry is such that the realized utility difference between x and y is often positive, generating $p > 1/2$, but takes large absolute values whenever it is negative. We will argue that asymmetric distributions in fact arise very naturally, even in conventional additive noise models where the utility of an option is given by $\tilde{u}(x) = u(x) + \tilde{\epsilon}(x)$. Wrongfully assuming symmetry then leads to false inferences about preferences.³

Now assume that we have data on the joint distribution of choices and response times. We will show in theorem 1 that $p > 1/2$ combined with, informally speaking, a comparatively slow choice of y relative to x is sufficient to conclude that the individual prefers x to y , even without making any assumptions about the shape of the utility distribution. A slow choice of y relative to x reveals that the utility difference cannot be distributed too asymmetrically in the way described above, because negative utility differences with large absolute values would generate quick choices of y on the basis of the chronometric relationship. This argument does not presume knowledge of the shape of the chronometric function beyond monotonicity (and some technical properties). More formally, let $F(x)(t)$ and $F(y)(t)$ be the cumulative distribution functions (CDFs) of response times conditional on the choice of x and y , respectively. Our criterion states that when we observe

$$F(y)(t) \leq \frac{p}{1-p} F(x)(t) \text{ for all } t \geq 0,$$

then any RUM with a chronometric function (RUM-CF) that rationalizes the data must satisfy $u(x) \geq u(y)$. A similar statement holds for strict preferences. In the limit as $p \rightarrow 1/2$, that is, as choice data alone becomes uninformative, this becomes the condition that choice of y must be slower than choice of x in the first-order stochastic dominance sense. As p grows, hence choice data becomes more indicative of a preference, the condition becomes weaker and requires only that choice of y is not much faster than choice of x .

We then study the case where the analyst has reasons to believe that utility differences are symmetrically distributed, as is often assumed in the

³ Our arguments are based on Horan, Manzini, and Mariotti (2019), who describe conditions under which differences of random utilities are symmetrically or asymmetrically distributed. Our example of an asymmetric distribution discussed in sec. III.A rests on independent, mean zero Gumbel errors like in any standard logit specification but with heteroskedasticity across the options. Given our limited understanding of how utility is constructed and how noise affects evaluation in preferential choice problems (which often involve trade-offs between multiple attributes), we see no good reasons for justifying the distributional assumptions required for symmetry, except tractability.

literature (e.g., in any application with logit or probit choice). It then follows immediately that $p > 1/2$ implies $u(x) > u(y)$, so preferences are revealed by choices without response times. But now we show that the use of response time data enables the identification of preferences for choice pairs outside the set of available choice data. For the case of deterministic choices and deterministic response times, this has been noted before. Krajbich, Oud, and Fehr (2014) argue that a slow choice of z over x combined with a quick choice of the same z over y reveals a preference for x over y , even though the choice between x and y is not directly observed and a transitivity argument is not applicable. The idea is that, based on the chronometric relationship, the positive utility difference $u(z) - u(x)$ must be smaller than the positive utility difference $u(z) - u(y)$, which implies $u(x) > u(y)$. To date, however, it has remained an open question how to implement this idea, since real-world choices and response times are stochastic, and hence it is unclear what “choice of z over x ” and “slow versus fast” exactly mean. For instance, is “faster than” defined in terms of mean response times, median response times, or some other characteristic of the response time distribution? Our theorem 2 provides an answer to that question. Suppose that z is chosen over x with a probability $p(z, x) > 1/2$, which indeed implies $u(z) - u(x) > 0$ in the symmetric noise case. Then we define $\theta(z, x)$ as a specific percentile of the response time distribution for z , namely, the $0.5/p(z, x)$ percentile. Analogously, if z is chosen over y with a probability $p(z, y) > 1/2$, which implies $u(z) - u(y) > 0$, the corresponding percentile $\theta(z, y)$ can be defined. Our result shows that these observable percentiles are the appropriate measure of preference intensity for the stochastic setting, in the sense that $\theta(z, x) > \theta(z, y)$ implies $u(z) - u(x) < u(z) - u(y)$ and hence a revealed (strict) preference for x over y . That is, inference cannot be based on mean, median, maximum, or minimum response times. The correct measurement is the $0.5/p$ percentile of the distribution of response times, which requires using a different percentile for each choice pair, adjusting for the respective choice frequencies. Since a revealed strict preference for x over y translates into a choice probability $p(x, y) > 1/2$ when the utility distribution is symmetric, this quantification generates out-of-sample predictions that are easy to test empirically.

In the traditional approach without response times, making out-of-sample predictions requires even stronger distributional assumptions than just symmetry. RUMs like probit or logit are instances of Fechnerian models (Debreu 1958; Moffatt 2015), in which the utility difference between the two options follows the exact same distributional form in all binary choice problems. With this Fechnerian assumption (but without assuming a specific functional form for the distribution), already the choice observation $p(z, x) < p(z, y)$ reveals a preference for x over y . Put differently, the Fechnerian assumption enables an exhaustive elicitation of ordinal preferences outside the data set. However, the use of response time data makes it

now even possible to move beyond ordinal preferences and make predictions of precise choice probabilities. Theorem 3 provides a closed-form formula to predict $p(x, y)$ on the basis of observables—that is, choice probabilities and response times—from only the binary choices between z and x and between z and y .

The general pattern that emerges from our results is that response time data allow us to obtain results that would otherwise require an additional distributional assumption, which might be empirically unjustified. Response time data make it possible to get rid of assumptions because the distribution of response times contains information about the distribution of utility noise. This enables the revelation of preferences without any distributional assumptions, makes it possible to extrapolate preferences to cases for which no choice data exist with a symmetry assumption, and even generates precise probability predictions with the Fechnerian assumption.

Our theorem 1 provides a robust sufficient condition for preference revelation, which essentially goes from data to models. To investigate how much bite our criterion has, we also look at the converse implication from models to data; that is, we study stochastic choice functions with response times (SCF-RTs) that are generated by standard models from the received literature. We take a specific data-generating process as given and apply our agnostic method that does not presume knowledge of the process to the resulting data set. We do this first for the whole class of RUM-CFs that have symmetric distributions, which contains the probit and logit models as special cases but goes far beyond them. We show that our criterion recovers all preferences correctly when any such model generated the data (proposition 4). In other words, our sufficient condition is also necessary and has maximal bite for the entire class of SCF-RTs generated by symmetric RUM-CFs. Even the analyst who believes in the probit or logit distribution can work with our criterion, because it must always hold in his data. Conversely, a data set where the condition in theorem 1 is violated cannot be explained through a symmetric RUM-CF. We then show that the result still holds with additional noise in response times, as long as the noise is from an independent source, like a stochastic chronometric function or imperfect observation, and does not systematically reverse the chronometric relationship (proposition 5).

Second, we study the class of drift-diffusion models (DDMs) with constant or collapsing decision boundaries, which are prominent in psychology and neuroscience (e.g., Ratcliff 1978; Shadlen and Kiani 2013). These models have recently attracted attention in economics because they can be derived from optimal evidence accumulation mechanisms (Drugowitsch et al. 2012; Tajima, Drugowitsch, and Pouget 2016; Fudenberg, Strack, and Strzalecki 2018; Baldassi et al. 2020). We show that our criterion again recovers all preferences correctly from data generated by a DDM (proposition 6). Hence, our previous statement on believers in

probit or logit models also applies to believers in DDMs. Again, as a consequence of this result, a data set where the condition in theorem 1 is violated cannot be explained through a DDM.

Finally, we apply our tools to an experimental data set from Clithero (2018). In that experiment, subjects made repeated choices between various snack food items, and response times were recorded. Our results are easy to apply by using nonparametric kernel density estimates of the response time distributions. We show that our condition from theorem 1 is fulfilled in 61% of all decision problems where choice was random. Hence, a preference is revealed without distributional assumptions in a majority of cases. In addition, the fact that in 39% of the cases the condition from theorem 1 is violated suggests that, in these cases, models with symmetric noise, including Fechnerian models as the standard logit and probit approaches, are not consistent with the data. These numbers illustrate the empirical relevance of our criterion but also warn against the unquestioned use of the symmetry assumption.

We then apply theorem 2 to predict choices in a second phase of the experiment. The accuracy of our out-of-sample predictions, which rest on no distributional assumptions other than symmetry, is remarkable. The prediction is correct in 80.7% of the cases, which is significantly better than for the logit model and indistinguishable from a computation-intensive DDM (Clithero 2018). We conduct an analogous analysis using theorem 3 to predict precise choice probabilities, assuming a Fechnerian structure of the utility noise. We again achieve a high prediction accuracy, significantly better than for the logit model. While the DDM estimated by Clithero (2018) performs even better in that case, our nonparametric method yields its high accuracy in a straightforward way and does not require structural estimation.

Our paper is related to the vast empirical and theoretical literature on RUMs, which have a long history going back to Thurstone (1927), Luce (1959), and Marschak (1960). These models became more widely used in economics after McFadden (1974; see also McFadden 2001). In the meantime, they have become a dominant workhorse in applied economics, as indicated by the more than 19,000 Google Scholar citations to date of McFadden (1974). We believe that our paper solves a fundamental issue in this approach by showing how the rigorous and principled use of response times greatly improves the ability to identify preferences. More generally, our paper illustrates how the use of nonchoice data—such as data on physiological, neural, and attentional processes that carry information about preferences—may help solving important problems in economics.

In addition, our paper contributes to a large literature in psychology, neuroeconomics, and neuroscience on response times and process models (e.g., Ratcliff and Rouder 1998; Krajbich, Armel, and Rangel 2010; Krajbich and Rangel 2011; Ratcliff et al. 2016). A standard view of human

decision-making in neuroscience (e.g., Shadlen and Kiani 2013) is that the chronometric function reflects neural processes of value comparison between the options and that these processes take longer to differentiate closer values. Our proposition 6 offers a test of whether DDMs can be the data-generating process for a given data set. In addition, our theorem 1 radically simplifies preference identification relative to structural estimations of DDMs.

Our paper is also related to the emerging literature on response times and evidence accumulation models in economics (Woodford 2014; Clithero 2018; Fudenberg, Strack, and Strzalecki 2018; Kononov and Krajbich 2019; Baldassi et al. 2020; Schotter and Trevino 2021). The chronometric relationship between response times and utility differences, in particular, has only recently become an object of interest in economics (e.g., Krajbich et al. 2015; Alós-Ferrer et al. 2016; Echenique and Saito 2017). The chronometric relationship has been formalized by Fudenberg, Strack, and Strzalecki (2018), who show that even in an optimal dynamic process of evidence accumulation, it takes longer to discover that one is close to indifference than to recognize a strong preference (see also Woodford 2014; for an axiomatization of evidence accumulation models, see Baldassi et al. 2020). Another possible interpretation of the chronometric phenomenon is that it captures an inherent suboptimality of human decision-making. In line with this interpretation, Krajbich, Oud, and Fehr (2014) show that choices can sometimes be improved by interventions that force people to spend less time on single decisions. Which of these interpretations one favors is inconsequential for our analysis.

Our approach is suitable for the analysis of decisions where an internal preference exists and has to be discovered by the decision maker. In our view, these decisions make up a large part of our daily choices. It is less suitable for life-changing decisions where a complex problem has to be solved, such as whether to accept a job or the parents' choice of the right school for their children. Such decisions take a long time not primarily because the options are similar but because they are enormously complex and multidimensional. Of course, even our daily choices vary in complexity. The choice between two snack food items is different from the choice between two means of transportation, and therefore we should expect to see different response times even when a similar utility difference is involved. Our approach assumes that the same chronometric function applies across the different binary choice problems, and therefore it can be used whenever the options are sufficiently similar in complexity to warrant this assumption. For instance, we would be wary to apply our methodology when some options are complex lotteries while others are simple deterministic payoffs or to a framework where a known heuristic or cognitive bias might affect some decisions but not others. The same reservations apply to conventional RUMs without response

times, which require that the same psychometric function is valid across choice problems (see, e.g., He and Natenzon 2018). In fact, as we will argue below, our definition of a RUM is more general than conventional ones, so that in principle it even allows us to capture certain pair-specific complexity differences across the binary choice problems.

The paper is organized as follows. Section II presents the formal setting. Section III develops the main results, devoting separate subsections to the unrestricted, symmetric, and Fechnerian cases. Section IV shows that choice data generated by standard models from economics and psychology fulfill our main criterion for preference revelation. Section V contains our empirical application. Section VI discusses the related literature in more detail, and section VII concludes. All proofs omitted from the main text can be found in the appendix (apps. A–C are available online).

II. Formal Setting and Definitions

Let X be a finite set of options. Denote by $C = \{(x, y) | x, y \in X, x \neq y\}$ the set of all binary choice problems, so (x, y) and (y, x) both represent the problem of choice between x and y . Let $D \subseteq C$ be the set of choice problems on which we have data, assumed to be nonempty and symmetric; that is, $(x, y) \in D$ implies $(y, x) \in D$. To economize notation, we let the set D be fixed throughout.

DEFINITION 1. A *stochastic choice function* (SCF) is a function p assigning to each $(x, y) \in D$ a probability $p(x, y) > 0$, with the property that $p(x, y) + p(y, x) = 1$.

In an SCF, $p(x, y)$ is interpreted as the probability of choosing x in the binary choice between x and y , and $p(y, x)$ is the probability of choosing y . The assumption that $p(x, y) > 0$ for all $(x, y) \in D$ implies that choice is stochastic in a nondegenerate sense, because each alternative is chosen with strictly positive probability.

Since there is no universally agreed-upon definition of RUMs, we will work with a fairly general definition that encompasses several previous ones. In particular, it is convenient for our analysis to directly describe for each $(x, y) \in C$ the distribution of the utility difference between the two options.

DEFINITION 2. A *random utility model* (RUM) is a pair $(u, \tilde{v})$ where $u: X \rightarrow \mathbb{R}$ 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: $\tilde{v}(x, y) = -\tilde{v}(y, x)$; and
- RUM.3: The support of $\tilde{v}(x, y)$ is connected.

In a RUM, the utility function u represents the underlying preferences that the analyst aims to uncover, while the random variables $\tilde{v}(x, y)$ also incorporate the noise, modeled directly as random pairwise utility differences. That is, the density $g(x, y)$ describes the distribution of the random utility difference between x and y . The noise is assumed to have zero mean, so the expected value of the random utility difference $\tilde{v}(x, y)$ must be $u(x) - u(y)$, as required by RUM.1. This condition can be spelled out in terms of the density $g(x, y)$ as

$$\int_{-\infty}^{+\infty} v g(x, y)(v) dv = u(x) - u(y).$$

We will also use the notation $v(x, y) = u(x) - u(y)$ for $\mathbb{E}[\tilde{v}(x, y)]$. Condition RUM.2 states that $\tilde{v}(x, y)$ and $\tilde{v}(y, x)$ describe the same random utility difference but with opposite signs; that is, $g(x, y)(v) = g(y, x)(-v)$ for all $v \in \mathbb{R}$. Finally, RUM.3 is a regularity condition stating that there are no gaps in the distribution of a pair's utility differences.

Our definition reflects the conventional idea that RUMs consist of a deterministic utility function plus mean zero noise terms, as typically implemented in discrete choice and microeconometrics (see Ben-Akiva and Lerman 1985; for a historical account, see McFadden 2001). Those approaches define a RUM by random utilities $\tilde{u}(x) = u(x) + \tilde{\epsilon}(x)$ for each option. Equivalently, a RUM of this kind can be given directly by a vector $\tilde{u} = (\tilde{u}(x))_{x \in X}$ of random variables, with the underlying deterministic utility function u being defined through $u(x) = \mathbb{E}[\tilde{u}(x)]$. Those models are particular cases of ours, taking $\tilde{v}(x, y) = \tilde{u}(x) - \tilde{u}(y)$. Our approach is more general because the utility differences $\tilde{v}(x, y)$ are unrestricted across choice pairs; that is, the realized utility difference is not constrained to be a difference of realized utilities. This allows us to accommodate pair-specific factors other than utility differences that may affect choice probabilities (and response times), such as dominance relations between some options (see He and Natenzon 2018 and our discussion in sec. VI.C). An additional benefit of this generality is that our positive results on preference revelation become stronger, because they hold within a larger class of models. The added generality does not matter for the impossibility result in proposition 1, which would also hold for RUMs of the form $\tilde{u}(x) = u(x) + \tilde{\epsilon}(x)$.

A RUM generates choices by assuming that the option chosen is the one with the larger realized utility.

DEFINITION 3. A RUM $(u, \tilde{v})$ rationalizes an SCF p if $p(x, y) = \text{Prob}[\tilde{v}(x, y) > 0]$ holds for all $(x, y) \in D$.

For the particular case where $\tilde{v}(x, y) = \tilde{u}(x) - \tilde{u}(y)$, the definition above just specifies $p(x, y) = \text{Prob}[\tilde{u}(x) > \tilde{u}(y)]$. Note that since $\tilde{v}(x, y)$ is

assumed to have a density $g(x, y)$, whether one writes $\text{Prob}[\tilde{v}(x, y) > 0]$ or $\text{Prob}[\tilde{v}(x, y) \geq 0]$ is inconsequential. Denote the CDF derived from $g(x, y)$ by $G(x, y)$. Then, the probability that the utility of y exceeds that of x is $G(x, y)(0)$. As a consequence, definition 3 can be alternatively stated as the condition that

$$G(x, y)(0) = p(y, x)$$

for all $(x, y) \in D$.

We now extend the framework and include response times by adding conditional response time distributions for each choice. This is the easiest way of describing a joint distribution over choices and response times.

DEFINITION 4. A *stochastic choice function with response times* (SCF-RT) is a pair (p, f) where p is an SCF and f assigns to each $(x, y) \in D$ a strictly positive density function $f(x, y)$ on $\mathbb{R}_+$.

The density $f(x, y)$ describes the distribution of response times conditional on x being chosen in the binary choice between x and y . The corresponding CDF is denoted by $F(x, y)$. It would be straightforward to introduce lower or upper bounds on response times, for instance, because of a nondecision time or a maximal observed response time. We refrain from doing so here for notational convenience and comparability with the literature (e.g., Fudenberg, Strack, and Strzalecki 2018).

DEFINITION 5. A *random utility model with a chronometric function* (RUM-CF) is a triple $(u, \tilde{v}, r)$ where $(u, \tilde{v})$ is a RUM and $r: \mathbb{R}_{++} \rightarrow \mathbb{R}_+$ is a continuous function that is strictly decreasing in v whenever $r(v) > 0$, with $\lim_{v \rightarrow 0} r(v) = +\infty$ and $\lim_{v \rightarrow \infty} r(v) = 0$.

In a RUM-CF, r represents the chronometric function. It maps realized utility differences v into response times $r(|v|)$, such that larger absolute utility differences generate shorter response times. The assumption that $\lim_{v \rightarrow 0} r(v) = +\infty$ and $\lim_{v \rightarrow \infty} r(v) = 0$ ensures that the model can encompass all response times observed in an SCF-RT. Our definition allows for functions like $r(v) = 1/v$ that are strictly decreasing throughout and also for functions that reach $r(v) = 0$ for large enough v . The latter case will arise when we construct chronometric functions from sequential sampling models in section IV. Figure 1 illustrates both cases, taking advantage of the fact that the inverse $r^{-1}(t)$ is well defined for the restriction of r to the subset where $r(v) > 0$.

In addition to choices, a RUM-CF generates response times by assuming that the realized response time is related to the realized utility difference through function r . Specifically, given a RUM-CF $(u, \tilde{v}, r)$ and a pair $(x, y) \in C$, the random variable describing the response times when x is chosen over y is given by

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

conditional on $\tilde{v}(x, y) > 0$. This motivates the following definition.

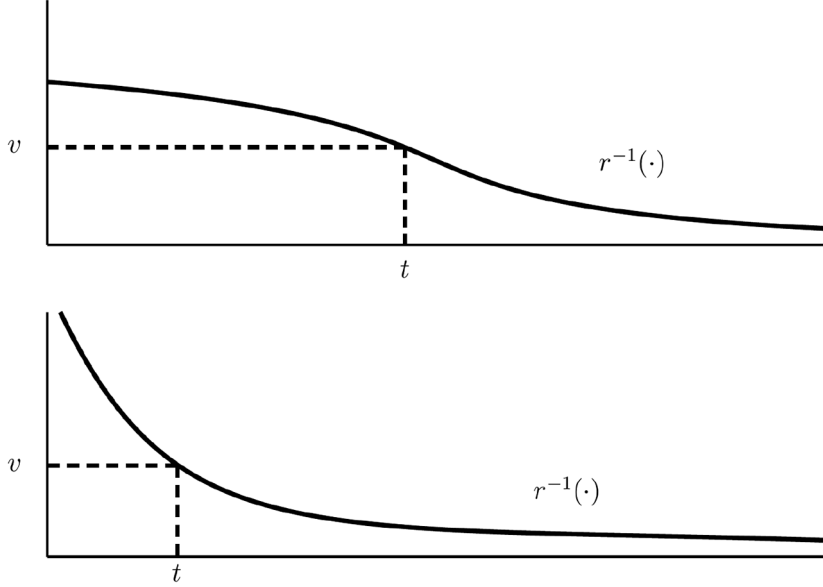


FIG. 1.—Two illustrations of chronometric functions r , mapping realized utility differences (vertical axis) into response times (horizontal axis).

DEFINITION 6. A RUM-CF $(u, \tilde{v}, r)$ rationalizes an SCF-RT (p, f) if $(u, \tilde{v})$ rationalizes p and $F(x, y)(t) = \text{Prob}[\tilde{t}(x, y) \leq t \mid \tilde{v}(x, y) > 0]$ holds for all $t > 0$ and all $(x, y) \in D$.

The probability of a response time of at most t , conditional on x being chosen over y , is the probability that the realized utility difference is at least $r^{-1}(t)$, conditional on that difference being positive. Since this probability can be calculated as $[1 - G(x, y)(r^{-1}(t))]/[1 - G(x, y)(0)]$, the condition in definition 6 is equivalent to

$$F(x, y)(t) = \frac{1 - G(x, y)(r^{-1}(t))}{1 - G(x, y)(0)} \quad (1)$$

for all $t > 0$ and all $(x, y) \in D$.

An alternative approach would have been to assume that response time is a decreasing function of the true absolute utility difference $|v(x, y)|$ between the two options as opposed to the realized noisy ones. A first drawback of this approach is that response times would be predicted to be deterministic, in contradiction to all available evidence. Hence, a second source of noise would have to be introduced, for instance, by making

the chronometric function stochastic. Without additional ad hoc assumptions, any such model would predict that the conditional distributions of response times for each of the two choices are identical, a further prediction not borne out by the data. Our approach is more parsimonious, as it requires only one source of randomness (in utility) to generate both stochastic choices and stochastic response times, and it does not make the implausible prediction of independence between choices and response times.⁴ A second drawback of the alternative approach is that the response time for two identical options would be predicted to be infinite, which is unreasonable. In contrast, our model does not make such a prediction, because even with indifference $v(x, y) = 0$, the stochastic utility difference $\tilde{v}(x, y)$ has mean zero but is different from zero with probability 1, generating continuous distributions of response times.⁵

When studying stochastic choice data, the analyst might be interested in or willing to make specific assumptions about the distribution of random utility. In doing so, the analyst accepts a restriction to a specific subclass of RUMs. Those might range from symmetry of each density $g(x, y)$ to specific functional forms. We say that an SCF (SCF-RT) is rationalizable within some class of models if there exists a RUM (RUM-CF) in that class that rationalizes it.

DEFINITION 7. Within a class of models, a rationalizable SCF (SCF-RT) reveals a preference for x over y if all RUMs (RUM-CFs) in the class that rationalize it satisfy $u(x) \geq u(y)$. It reveals a strict preference for x over y if all RUMs (RUM-CFs) in the class that rationalize it satisfy $u(x) > u(y)$.

III. Revealed Preference

In this section, we investigate the use of response times for preference revelation. Specifically, we are interested in preference revelation within difference classes of RUMs and how the addition of response times improves the results.

A. The Unrestricted Case

The first observation is that, without further restrictions on the utility distributions and without the use of response times, nothing can be

⁴ That said, a second source of independent noise, like a stochastic chronometric function, can be introduced without changing our main insights, as we will show in sec. IV.

⁵ A realized stochastic utility difference of zero would give rise to an infinite response time, but this happens with probability zero. The generated distributions can have finite means, as we would expect in data. Our model does not even predict that these mean response times must be decreasing in the underlying deterministic utility difference. That is an implication with additional assumptions fulfilled by standard models with logit or probit errors but does not necessarily hold in the general case.

learned from choice probabilities. This is well known among specialists in stochastic choice theory, and hence we do not claim originality.⁶

PROPOSITION 1. Within the class of all RUMs, a rationalizable SCF reveals no preference between any x and y , with $x \neq y$.

The intuition for the result is simple. Data on the choice between x and y allow us to learn the value $G(x, y)(0)$, but without distributional assumptions, this does not tell us whether the expected value $v(x, y) = u(x) - u(y)$ is positive or negative, which is what we are interested in.

A solution to the problem would be to impose the seemingly innocuous assumption of symmetry of the distribution. In that case, $G(x, y)(0) \leq 1/2$ indeed implies $v(x, y) \geq 0$, and $G(x, y)(0) \geq 1/2$ implies $v(x, y) \leq 0$. We will investigate the symmetry assumption and the scope for response times to improve preference revelation under that assumption in section III.B. First, however, we want to illustrate with a simple example that symmetry sometimes may not be an innocuous assumption at all.

Example 1.—Consider a simple RUM with additive noise, where $\tilde{u}(x) = u(x) + \tilde{\epsilon}(x)$ and $\tilde{u}(y) = u(y) + \tilde{\epsilon}(y)$. The noise terms $\tilde{\epsilon}(x)$ and $\tilde{\epsilon}(y)$ are independent and have zero mean. It is often assumed in applied work that they follow an identical Gumbel distribution, with CDF

$$H(z) = e^{-e^{-(z/\beta + \gamma)}},$$

where $\beta > 0$ is a parameter and $\gamma \approx 0.5772$ is the Euler-Mascheroni constant. The variance of this distribution is $\beta^2 \pi^2 / 6$. The difference $\tilde{v}(x, y) = \tilde{u}(x) - \tilde{u}(y)$ then follows a logistic distribution, which is easy to describe in closed form and generates the well-known logit choice probabilities (see also our discussion in sec. IV). This distribution is indeed symmetric around its mean $v(x, y) = u(x) - u(y)$.

Horan, Manzini, and Mariotti (2019) investigate the case where $\tilde{\epsilon}(x)$ and $\tilde{\epsilon}(y)$ are still independent, mean zero Gumbel random variables but with different variances. If the noise terms capture mistakes in the evaluation of the options, it is plausible that some options are more difficult to evaluate than others, resulting in option-specific variances. If the noise terms reflect random variation in the tastes of a given individual or heterogeneity in a population of agents, there is no reason to assume identical distributions for different options either. However, the distribution of $\tilde{v}(x, y)$ is no longer symmetric for option-specific parameters β_x and β_y of the underlying Gumbel distributions. Unfortunately, that distribution cannot be described in closed form when $\beta_x \neq \beta_y$ (presumably one of the reasons why the literature has focused on the knife-edge case where $\beta_x = \beta_y$).

⁶ See, e.g., the discussion in Haile, Hortaçsu, and Kosenok (2008). A similar problem arises in ordered response models that are commonly used to analyze ordinal survey data (see Bond and Lang 2019).

The following numerical example is inspired by example 3 in Horan, Manzini, and Mariotti (2019). Suppose that $u(x) = 1$ and $u(y) = 3/4$, so $v(x, y) = 1/4 > 0$ and the individual strictly prefers x over y . Suppose furthermore that $\beta_x = 1$ and $\beta_y = 2$, for instance, because y is more difficult to evaluate than x , and thus its utility is more noisy. Then, it follows that the probability of choosing x is $p(x, y) = \text{Prob}[\tilde{u}(x) > \tilde{u}(y)] \approx 0.49 < 1/2$.

Consider now an analyst who does not know the data-generating process. Suppose that this analyst erroneously imposes the symmetry assumption when analyzing the SCF with the intention of uncovering the underlying preference. This analyst will correctly deduce that $G(x, y)(0) \approx 0.51$ but then, applying symmetry, incorrectly conclude that $v(x, y)$ is strictly negative, that is, that the SCF reveals a strict preference for y over x .

This example is meant to illustrate (1) that symmetry is not necessarily a plausible restriction and (2) that erroneously making the symmetry assumption can lead to wrong inferences about preferences.⁷

The following result shows that if response times are available, it may be possible to learn a preference even in the unrestricted class of models. We first introduce the following new concepts. Given two CDFs G and H on $\mathbb{R}_+$ and a constant $q \geq 1$, we say that G q -first-order stochastically dominates H (also written G q -FSD H) if

$$G(t) \leq q \cdot H(t) \text{ for all } t \geq 0.$$

If, additionally, the inequality is strict for some t , then G strictly q -first-order stochastically dominates H (written G q -SFSD H). For $q = 1$, these concepts coincide with the standard notions of first-order stochastic dominance. They are weaker requirements when $q > 1$ and possibly substantially so, because the dominating function G can lie above H to an extent constrained only by the ratio q . In particular, q -FSD implies q' -FSD whenever $q \leq q'$. Furthermore, for any two distributions G and H for which $G(t)/H(t)$ is bounded, we can always find a large enough q such that G q -FSD H .

THEOREM 1. Within the class of all RUM-CFs, a rationalizable SCF-RT reveals a preference for x over y if $F(y, x)$ q -FSD $F(x, y)$ and a strict preference if $F(y, x)$ q -SFSD $F(x, y)$ for $q = p(x, y)/p(y, x)$.

Proof.—Let $(u, \tilde{v}, r)$ be any RUM-CF that rationalizes an SCF-RT (p, f) , and consider any $(x, y) \in D$. By (1), it holds that

$$1 - G(x, y)(r^{-1}(t)) = p(x, y)F(x, y)(t) \quad (2)$$

for all $t > 0$. Since RUM.2 implies $1 - G(y, x)(v) = G(x, y)(-v)$, for $(y, x) \in D$ we analogously obtain

⁷ Other natural examples with asymmetric utility difference distributions can also be constructed in the class of random parameter models, where the decision maker's preference parameters are stochastic (Apesteguía and Ballester 2018).

$$G(x, y)(-r^{-1}(t)) = p(y, x)F(y, x)(t) \quad (3)$$

for all $t > 0$. With the definition of $Q(x, y)(t) = (p(x, y)F(x, y)(t)) / (p(y, x)F(y, x)(t))$, we therefore have

$$Q(x, y)(t) = \frac{1 - G(x, y)(r^{-1}(t))}{G(x, y)(-r^{-1}(t))} \quad (4)$$

for all $t > 0$.

Now suppose that $F(y, x)$ q -FSD $F(x, y)$ for $q = p(x, y)/p(y, x)$. This can equivalently be written as $Q(x, y)(t) \geq 1$ for all $t > 0$. Hence, it follows from (4) that

$$G(x, y)(-r^{-1}(t)) \leq 1 - G(x, y)(r^{-1}(t))$$

for all $t > 0$. We claim that this implies

$$G(x, y)(-v) \leq 1 - G(x, y)(v) \quad (5)$$

for all $v \geq 0$. The inequality follows immediately for any v for which there exists $t > 0$ such that $r^{-1}(t) = v$. For $v = 0$, it follows from continuity of $G(x, y)(v)$. For any v with $r(v) = 0$, it follows because in that case $G(x, y)(v) = 1$ and $G(x, y)(-v) = 0$, as otherwise the RUM-CF would generate an atom at the response time of zero.

Define a function $H : \mathbb{R} \rightarrow [0, 1]$ by

$$H(v) = \begin{cases} 1 - G(x, y)(-v) & \text{if } v \geq 0, \\ G(x, y)(v) & \text{if } v < 0. \end{cases}$$

Observe that H is the CDF for a distribution that is symmetric around zero and continuous except (possibly) for an atom at zero. Hence,

$$\begin{aligned} \int_{-\infty}^{+\infty} v dH(v) &= \int_{(-\infty, 0)} v dH(v) + \int_{(0, +\infty)} v dH(v) \\ &= \int_{-\infty}^0 v g(x, y)(v) dv + \int_0^{+\infty} v g(x, y)(-v) dv \\ &= \int_{-\infty}^0 v g(x, y)(v) dv - \int_{-\infty}^0 v g(x, y)(v) dv = 0. \end{aligned}$$

Observe furthermore that (5) implies $G(x, y)$ 1-FSD H . Hence, we have

$$v(x, y) = \int_{-\infty}^{+\infty} v dG(x, y)(v) \geq \int_{-\infty}^{+\infty} v dH(v) = 0, \quad (6)$$

that is, a revealed preference for x over y .

If $F(y, x)$ q -SFSD $F(x, y)$ for $q = p(x, y)/p(y, x)$, then (5) is strict for some $v \geq 0$. Hence, $G(x, y)$ 1-SFSD H and the inequality in (6) is strict, that is, a revealed strict preference for x over y . QED

The basic idea behind theorem 1 is that the observable distributions of response times provide information about the unobservable distributions of utilities, based on the chronometric relationship.

To understand the precise condition, assume first that $q = p(x, y)/p(y, x) = 1$ for some $(x, y) \in D$; that is, both options are equally likely to be chosen. Any RUM-CF that rationalizes this choice must satisfy $G(x, y)(0) = 1/2$. Furthermore, note that the distribution of $\tilde{v}(x, y)$ conditional on $\tilde{v}(x, y) > 0$ generates $F(x, y)$, and the distribution of $\tilde{v}(x, y)$ conditional on $\tilde{v}(x, y) < 0$ generates $F(y, x)$. Thus, if we additionally observe that $F(x, y)(t) = F(y, x)(t)$ for all $t \geq 0$ (i.e., identical response time distributions for the two options), then we can conclude that the shape of the utility difference distribution must be identical on the positive and on the negative domain. This requires no knowledge of the properties of r beyond monotonicity. Hence, we have verified that the distribution is symmetric around zero, so its mean $v(x, y)$ is zero. Our theorem indeed implies a revealed preference for x over y and for y over x in this case, which we also call a *revealed indifference between x and y* . If, by contrast, we observe that $F(y, x)$ 1-SFSD $F(x, y)$ (i.e., the choice of y is systematically slower than the choice of x), we can conclude that the utility difference distribution is asymmetric and takes systematically larger absolute values on the positive than on the negative domain. Hence, its mean $v(x, y)$ is strictly larger than zero, which translates into a revealed strict preference for x over y . Finally, if we observe $q = p(x, y)/p(y, x) > 1$, then to obtain a revealed preference for x over y , it is sufficient that the choice of y is not too much faster than the choice of x , as captured by our concept of q -first-order stochastic dominance. If choice behavior is already indicative of a particular preference, then the response time distributions just need to confirm that the utility difference distribution is not strongly asymmetric in the reverse direction.⁸

Remark.—We have stated theorem 1 as providing a sufficient condition for preference revelation. In sections IV and V, we will illustrate that this sufficient condition holds often and is hence useful. However, we could have chosen a different formal approach, casting the condition as an axiom and formulating the analysis in terms of data sets compatible with that axiom. Several results in section IV will show that entire subclasses

⁸ Theorem 1 and our further results focus on uncovering the sign of the mean of $\tilde{v}(x, y)$, because this mean equals $u(x) - u(y)$ and therefore informs about the ordinal preferences represented by u for either normative or positive reasons. Potentially, one could be interested in uncovering also other summary statistics of $\tilde{v}(x, y)$, and our tools may be helpful for that purpose, but the relevance of other statistics is not obvious from the point of view of revealed preference theory.

of models (symmetric RUM-CFs and DDMs) would be covered with this approach. We prefer to state theorem 1 as we do because we would not see a violation of the q -FSD condition as a fundamental problem for our approach. For instance, in RUM-CFs with asymmetric noise, it is possible that a given strict preference does not generate data fulfilling our sufficient condition. Therefore, a violation of the sufficient condition contradicts specific subclasses of models but not the general approach.

In the remainder of the section, we will discuss several implications of theorem 1. First, the result can be extended by completing the revealed preferences in a transitive way. A formal statement and proof of this claim can be found in appendix B.

Second, a large literature in psychology has examined the question of whether, in tasks where errors can be identified objectively, those are faster or slower than correct responses. Although in many situations errors tend to be slower on average than correct responses, this is not always the case, and no general conclusion can be drawn.⁹ Our result might help to clarify this relation for preferential choice. While the definition of error and correct response is not obvious *ex ante* for preferential choice, *ex post* we can call the choice of y a *revealed error* and the choice of x a *revealed correct response* when x is revealed to be strictly preferred over y . Translated into this language, it follows immediately that slow choices in the first-order stochastic dominance sense indeed reveal an error, provided that choice probabilities are at least minimally informative. However, our actual condition is weaker than first-order stochastic dominance, and hence it might be fulfilled even in cases where errors are faster on average. Thus, our approach is in principle compatible with the seemingly puzzling evidence from psychology.

Third, we have described SCF-RTs by unconditional choice probabilities and conditional response time distributions for each choice. This is the natural extension of SCFs and allowed us to work out the intuition for our result. Alternatively, we could have described the joint distribution over choices and response times by an unconditional response time distribution and conditional choice probabilities for each response time. Let $P(x, y)(t)$ denote the probability of a choice of x over y conditional on choice taking place before time t . The ratio of these probabilities can be calculated as

⁹ For an overview of classical results, see Luce (1986, sec. 6.4.3). The picture is further complicated if decisions are subject to extraneous impulsive tendencies, such as, e.g., alternative decision processes reflecting underlying biases. For instance, assumptions linking response times to whether an answer is more intuitive or more deliberative are common in the literature on dual-process thinking (e.g., Kahneman 2003), but as pointed out by Krajbich et al. (2015), apparent results in this direction might sometimes hide chronometric effects. Additionally, Achtziger and Alós-Ferrer (2014) show that a simple dual-process model predicts that errors might be either faster or slower than correct responses, depending on whether the underlying processes are in conflict or aligned.

$$Q(x, y)(t) = \frac{P(x, y)(t)}{P(y, x)(t)} = \frac{p(x, y)F(x, y)(t)}{p(y, x)F(y, x)(t)}.$$

Hence, the condition that $F(y, x) \ q\text{-FSD} \ F(x, y)$ for $q = p(x, y)/p(y, x)$ in theorem 1 is equivalent to

$$Q(x, y)(t) \geq 1 \text{ for all } t > 0. \quad (7)$$

An analogous formulation holds for the strict case. For a revealed preference without distributional assumptions, we can thus also check if x is more likely to be chosen than y before all times t . The simple requirement $p(x, y) \geq p(y, x)$ obtains as a special case of this in the limit as $t \rightarrow \infty$.

The formulation based on $Q(x, y)(t)$ suggests a natural but substantially stronger condition. Let $p(x, y)(t)$ denote the probability of a choice of x over y conditional on choice taking place at time t (rather than before t). We obtain the ratio

$$q(x, y)(t) = \frac{p(x, y)(t)}{p(y, x)(t)} = \frac{p(x, y)f(x, y)(t)}{p(y, x)f(y, x)(t)}$$

and can state the following corollary to theorem 1.

COROLLARY 1. Within the class of all RUM-CFs, a rationalizable SCF-RT reveals a preference for x over y if $q(x, y)(t) \geq 1$ for almost all $t \geq 0$ and a strict preference if, additionally, the inequality is strict for a set of t with positive Lebesgue measure.

The condition that x is more likely to be chosen than y at almost all times t can be interpreted as a requirement of stochastic consistency across response times. This is clearly stronger than (7). Appendix C contains an example in which $p(x, y)(t) < p(y, x)(t)$ holds for an interval of response times, so corollary 1 is not applicable, but a strict preference for x over y is still revealed by theorem 1. Hence, our main criterion arrives at a conclusion even though behavior displays some stochastic inconsistency across response times.¹⁰

The results in this section are interesting for two main reasons. First, for the analyst who is reluctant to make distributional assumptions in the context of RUMs, theorem 1 provides a robust criterion for preference revelation. On the basis of observed response times, it is often possible to deduce preferences without such assumptions. The criterion may lead to an incomplete revelation of preferences (we will return to this issue in secs. IV and V), but it avoids making mistakes like those illustrated in

¹⁰ Notice a similarity to Bernheim and Rangel (2009), who require agreement of choices across choice sets or frames to obtain a revealed preference. A first difference is that we study stochastic choice and contemplate probabilistic agreement of choices across response times. A second difference is that our main criterion can reveal a preference even if there is no such agreement.

example 1. Second, our criterion may be able to arbitrate if choice behavior violates stochastic transitivity (Tversky 1969; Rieskamp, Busemeyer, and Mellers 2006; Tsetsos et al. 2016). For example, assume that we observe a stochastic choice cycle with $p(x, y) > 1/2$, $p(y, z) > 1/2$, and $p(z, x) > 1/2$. Such a cycle cannot be rationalized by any model with symmetric utility distributions, but it may be rationalizable by a model with asymmetric utility distributions. In that case, at most two of the three binary choices can reveal a preference, showing which part of the cycle reflects true preferences. A similar argument applies if choices are affected by framing and we observe $p^f(x, y) > p^f(y, x)$ under frame f but $p^{f'}(x, y) < p^{f'}(y, x)$ under frame f' . Again, our response time criterion may be able to detect which frame induces choices that are probabilistically more in line with the true preferences. In other words, since a choice probability above $1/2$ is not yet sufficient to reveal a preference, RUM-CFs can explain cyclic or frame-dependent choices but are inconsistent with data sets for which our condition reveals cyclic or frame-dependent preferences. If the latter were observed, our model would be falsified.

B. The Symmetric Case

The assumption of symmetry is often accepted in the literature. Formally, a RUM $(u, \tilde{v})$ or RUM-CF $(u, \tilde{v}, r)$ is *symmetric* if each random variable $\tilde{v}(x, y)$ follows a distribution that is symmetric around its mean $v(x, y)$; that is, if for each $(x, y) \in C$ and all $\delta \geq 0$,

$$g(x, y)(v(x, y) + \delta) = g(x, y)(v(x, y) - \delta).$$

In contrast to proposition 1, this assumption allows us to learn preferences from observed choice probabilities.

PROPOSITION 2. Within the class of symmetric RUMs, a rationalizable SCF reveals a preference for x over y if $p(x, y) \geq p(y, x)$ and a strict preference if $p(x, y) > p(y, x)$.

This result is simple and well known, and we include a proof in appendix A only for completeness.¹¹ It allows us to deduce either a strict preference or an indifference for each observed choice pair, and it can be extended by completing the revealed preferences in a transitive way (see app. B).

Note that every preference that can be learned with the help of response times without distributional assumptions can also be learned without

¹¹ Its first statement in the economics literature that we are aware of is Manski (1977), but an earlier, closely related statement for general stochastic choice can be found already in Edwards (1954).

response times at the price of making the symmetry assumption.¹² But even if one is willing to make the symmetry assumption, the addition of response times again improves what can be learned about preferences, as the following result will show. It is based on triangulating a preference indirectly through comparisons with a third option. For each $(x, y) \in D$ with $p(x, y) > p(y, x)$, define $\theta(x, y)$ as the $0.5/p(x, y)$ percentile of the response time distribution of x ; that is,

$$F(x, y)(\theta(x, y)) = \frac{0.5}{p(x, y)}.$$

The percentile $\theta(x, y) > 0$ combines information about choice probabilities and response times. It is larger than the median but converges to the median when $p(x, y) \rightarrow 1$. More generally, $\theta(x, y)$ becomes smaller as $p(x, y)$ becomes larger or as the choice of x becomes faster in the usual first-order stochastic dominance sense. Hence, a small value of $\theta(x, y)$ is indicative of a strong preference for x over y . Comparison of these percentiles can make it possible to learn preferences for unobserved pairs $(x, y) \in C \setminus D$ even when transitivity is void.

Theorem 2. Within the class of symmetric RUM-CFs, a rationalizable SCF-RT reveals a preference for x over y , where $(x, y) \in C \setminus D$, if there exists $z \in X$ such that $\theta(x, z) \leq \theta(y, z)$ or $\theta(z, x) \geq \theta(z, y)$ and a strict preference if $\theta(x, z) < \theta(y, z)$ or $\theta(z, x) > \theta(z, y)$.

Proof.—Let $(u, \tilde{v}, r)$ be any symmetric RUM-CF that rationalizes an SCF-RT (p, f) . We first claim that for any $(x, y) \in D$ with $p(x, y) > p(y, x)$, and hence $v(x, y) > 0$ by symmetry, it holds that $\theta(x, y) = r(v(x, y))$. To see the claim, note that from (1) we obtain

$$p(x, y)F(x, y)(t) = 1 - G(x, y)(r^{-1}(t))$$

for all $t > 0$. Evaluated at $t = r(v(x, y))$, which is well defined because $v(x, y) > 0$ and strictly positive because otherwise the RUM-CF would generate an atom at the response time of zero, this yields

$$p(x, y)F(x, y)(r(v(x, y))) = 1 - G(x, y)(v(x, y)). \quad (8)$$

The right-hand side of (8) equals 0.5 by symmetry. Hence, we obtain

¹² This is true because $p(x, y) \geq p(y, x)$ is a necessary condition for a revealed preference for x over y according to theorem 1. The statement holds for weak but not necessarily for strict preferences. We can indeed have a revealed strict preference for x over y according to theorem 1 and a revealed indifference according to proposition 2, in case $p(x, y) = p(y, x) = 1/2$ and $F(y, x)$ 1-SFSD $F(x, y)$. Any symmetric RUM that rationalizes such an SCF must have $v(x, y) = 0$. However, there is no symmetric RUM-CF that rationalizes the SCF-RT because of the asymmetric response times. All rationalizing RUM-CFs must be asymmetric and have $v(x, y) > 0$.

$$F(x, y)(r(v(x, y))) = \frac{0.5}{p(x, y)},$$

and by definition of $\theta(x, y)$, it follows that $\theta(x, y) = r(v(x, y))$, proving the claim.

Consider now any $(x, y) \in C \setminus D$ for which there exists $z \in X$ with $\theta(x, z) \leq \theta(y, z)$, and hence $0 < r(v(x, z)) \leq r(v(y, z))$ by the above claim. Since r is strictly decreasing in v whenever $r(v) > 0$, it follows that $u(x) - u(z) = v(x, z) \geq v(y, z) = u(y) - u(z)$ and hence $v(x, y) = u(x) - u(y) \geq 0$, that is, a revealed preference for x over y . If $\theta(x, z) < \theta(y, z)$, all the inequalities must be strict, so the revealed preference is strict. The case where $\theta(z, x) \geq \theta(z, y)$ or $\theta(z, x) > \theta(z, y)$ is analogous. QED

It has been observed before that response times can be used to infer preferences for unobserved choices. Krajbich, Oud, and Fehr (2014) argue that a slow choice of z over x combined with a quick choice of the same z over y reveals a preference for x over y , even though the choice between x and y is not directly observed and a transitivity argument is not applicable. On the basis of the chronometric relationship, the positive utility difference $u(z) - u(x)$ must be smaller than the positive utility difference $u(z) - u(y)$, which implies $u(x) > u(y)$. It remained unclear how to generalize the idea to a stochastic framework. Our theorem 2 answers this question. The condition $\theta(z, x) \geq \theta(z, y)$ is the appropriate formulation of a stochastic choice of z over x being slower than a stochastic choice of z over y . Of course, an analogous argument applies to a quick choice of x over z combined with a slow choice of y over z , as formalized by our alternative condition $\theta(x, z) \leq \theta(y, z)$. Also, not too surprisingly, the preferences revealed by theorem 2 can further be completed in a transitive way (see app. B).

Importantly, we need to compare specific percentiles of the response time distributions that depend on choice probabilities and not just mean, median, or maximum response times. Up-to-date, empirical applications of response times have overwhelmingly been based on the use of a particular statistic, stating that, say, median response times are shorter or larger for certain types of decisions. Our analysis shows that such an approach is but a first approximation that we can strongly improve upon. Focusing on a particular statistic like the median treats all observations equally and neglects fundamental information, namely, choice frequencies. As already in the case of theorem 1, our analysis combines information about frequencies and response times. The condition in theorem 2 might well dictate to compare the 0.85 percentile of the distribution of response times for one choice to the 0.55 percentile of the distribution for a different choice.

The results in this section enable first out-of-sample predictions. Consider an unobserved choice problem $(x, y) \in C \setminus D$. If, on the basis of theorem 2, the SCF-RT reveals a strict preference for x over y in the class of

symmetric RUM-CFs, then we predict that $p(x, y) > p(y, x)$, because each symmetric model with $v(x, y) > 0$ generates such choice probabilities. If the SCF-RT reveals an indifference between x and y , then we can even predict the precise probabilities $p(x, y) = p(y, x) = 1/2$. We will test these predictions in our empirical application in section V.

C. The Fechnerian Case

Microeconomic models of random utility assume even more structure. For instance, the prominent probit or logit models are special cases of Fechnerian models, which go back to the representation result by Debreu (1958). A RUM $(u, \tilde{v})$ or RUM-CF $(u, \tilde{v}, r)$ is Fechnerian if the distribution of each random variable $\tilde{v}(x, y)$ has the same symmetric shape, which is just shifted so that its expected value becomes $v(x, y)$. Formally, there exists a common density g that is symmetric around zero and has full support; that is, $g(\delta) = g(-\delta) > 0$ for all $\delta \geq 0$, such that for each $(x, y) \in C$ and all $v \in \mathbb{R}$,

$$g(x, y)(v) = g(v - v(x, y)).$$

This additional structure makes it possible to deduce preferences through comparison with a third option, relying only on choice probabilities.

PROPOSITION 3. Within the class of Fechnerian RUMs, a rationalizable SCF reveals a preference for x over y , where $(x, y) \in C \setminus D$, if there exists $z \in X$ such that $p(x, z) \geq p(y, z)$ and a strict preference if $p(x, z) > p(y, z)$.

As in the case of proposition 2, this result is well known, and we provide a short proof in appendix A only for completeness.¹³ The transitive closure extension can be found in appendix B.

By proposition 3, we obtain a revealed preference for an unobserved pair $(x, y) \in C \setminus D$ whenever $(x, z), (y, z) \in D$ for some third option z . Hence, imposing the Fechnerian assumption enables an exhaustive elicitation of ordinal preferences even outside the choice data set without the use of response times (provided that the assumption is valid). We now show that the use of response times makes it possible to move beyond ordinal preferences and make out-of-sample predictions of precise choice probabilities.

DEFINITION 8. Within a class of models, a rationalizable SCF-RT predicts choice probability $\bar{p}(x, y)$ for a nonobserved choice $(x, y) \in C \setminus D$ if all RUM-CFs in the class that rationalize it satisfy $\text{Prob}[\tilde{v}(x, y) > 0] = \bar{p}(x, y)$.

THEOREM 3. Within the class of Fechnerian RUM-CFs, a rationalizable SCF-RT predicts a choice probability for each $(x, y) \in C \setminus D$ for which there exists $z \in X$ with $(x, z), (y, z) \in D$. The prediction is

¹³ The argument can be traced back to Fechner (1860) and Thurstone (1927). Within economics, it has been spelled out by Ballinger and Wilcox (1997).

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

Proof.—Let $(u, \tilde{v}, r)$ be any Fechnerian RUM-CF that rationalizes an SCF-RT (p, f) . For any fixed $(x, y) \in C$, this particular RUM-CF predicts

$$p(x, y) = \text{Prob}[\tilde{v}(x, y) > 0] = G(y, x)(0) = G(v(x, y)). \quad (9)$$

Let $(x, y) \in C \setminus D$ and $z \in X$ such that $(x, z), (y, z) \in D$. We distinguish three cases.

Case 1: $p(y, z) > 1/2$. From the proof of theorem 2, we already know that $\theta(y, z) = r(v(y, z))$. From (1), we obtain

$$p(x, z)F(x, z)(t) = 1 - G(x, z)(r^{-1}(t))$$

for all $t > 0$. Hence, by RUM.2 and the Fechnerian assumption,

$$\begin{aligned} p(x, z)F(x, z)(t) &= G(z, x)(-r^{-1}(t)) = G(-r^{-1}(t) - v(z, x)) \\ &= G(v(x, z) - r^{-1}(t)) \end{aligned}$$

for all $t > 0$, which for $t = r(v(y, z))$ yields

$$p(x, z)F(x, z)(r(v(y, z))) = G(v(x, z) - v(y, z)) = G(v(x, y)).$$

Combined with (9) and the above expression for $\theta(y, z)$, this implies

$$p(x, y) = p(x, z)F(x, z)(\theta(y, z)),$$

which is the model-independent prediction $\bar{p}(x, y)$ given in the statement.

Case 2: $p(y, z) = 1/2$. It follows from proposition 2 that $v(y, z) = 0$. We obtain

$$p(x, y) = G(v(x, y)) = G(v(x, z) - v(y, z)) = G(v(x, z)) = p(x, z),$$

which is the model-independent prediction $\bar{p}(x, y)$ given in the statement.

Case 3: $p(y, z) < 1/2$. It follows from proposition 2 that $v(z, y) > 0$. Following the same steps as in case 1 but with reversed order for each pair of alternatives yields the model-independent prediction

$$\bar{p}(x, y) = 1 - \bar{p}(y, x) = 1 - p(z, x)F(z, x)(\theta(z, y)),$$

as given in the statement.

To understand the probability formula in the theorem, just consider the case where $p(x, z) > p(y, z) > 1/2$. Then $u(x) > u(y) > u(z)$ must

hold under the Fechnerian assumption, where the first inequality follows from proposition 3 and the second inequality follows from proposition 2. Hence, we can conclude that the unknown $p(x, y)$ must be strictly smaller than $p(x, z)$, because Fechnerian choice probabilities are strictly monotone in the underlying utility differences $v(\cdot, \cdot)$ across binary choice problems. The theorem now shows that a prediction for $p(x, y)$ can be obtained by multiplying the observed $p(x, z)$ with a discounting factor $F(x, z)(\theta(y, z))$. This factor is an observable, response time–based indicator of the relative position of $u(y)$ within the interval $[u(x), u(z)]$.

Combined with the Fechnerian assumption, the use of response times allows us to predict exact choice probabilities out of sample. Without response times, making such a prediction would require assuming a complete and specific functional form for the utility distribution.¹⁴ Hence, in analogy to our earlier results, response times again serve as a substitute for stronger distributional assumptions.¹⁵ Further, as was the case for theorems 1 and 2, theorem 3 extracts information from the data by balancing choice frequencies and response times.

IV. Behavioral Models from Economics and Psychology

Theorem 1 provided a sufficient condition for preference revelation without distributional assumptions. This analysis left open two questions. First, which SCF-RTs are rationalizable by RUM-CFs? Second, how tight is the sufficient condition? In this section, we try to answer these questions by studying SCF-RTs that are generated by specific behavioral models from the literature. We will do this first for standard RUMs from economics and microeconometrics to which we add chronometric functions and second for standard sequential sampling models from psychology and neuroscience.

A. *The View from Economics*

Specific RUMs are commonly used for microeconomic estimation using choice data. These models typically start from a utility function $u : X \rightarrow \mathbb{R}$

¹⁴ Some weaker inequality predictions of choice probabilities are already possible under an additive RUM structure without additional distributional assumptions (see, e.g., Cohen and Falmagne 1990).

¹⁵ For the analyst who is willing to make the strong distributional assumptions required by probit or logit models, response times have no additional value when the available data on choice are rich. However, the literature has shown that the use of response times can be valuable even in the context of logit or probit models when choice data are scarce (e.g., Clithero 2018; Konovalov and Krájčich 2019). Our paper differs from these studies by showing that response time data enable the recovery of preferences even when rich choice data cannot recover them, i.e., in the absence of untestable assumptions on utility noise.

and add an error term with zero mean to each option, such that the overall utility of $x \in X$ is given by a random variable

$$\tilde{u}(x) = u(x) + \tilde{\epsilon}(x).$$

As a next step, even more specific distributional assumptions are imposed. Two popular examples are the probit and the logit model.

In the probit model, the errors are assumed to be normally distributed and independently and identically distributed (i.i.d.) across the options. The distribution of the random utility difference $\tilde{u}(x, y) = \tilde{u}(x) - \tilde{u}(y)$ is then also normal and can be described by

$$G(x, y)(v) = \Phi\left(\frac{v - v(x, y)}{\sigma}\right),$$

where σ is a standard deviation parameter and Φ is the CDF of the standard normal distribution. This simple specification gives rise to a Fechnerian model. Generalizations allow for heteroscedasticity or correlation between the error terms of different options, in which case the parameter σ becomes choice set dependent, written $\sigma(x, y)$. Such a generalized model is no longer Fechnerian but still symmetric.

In the logit model, the errors are assumed to follow a Gumbel distribution, again i.i.d. across the options. In that case, the random utility difference follows a logistic distribution described by

$$G(x, y)(v) = \left\{1 + e^{-[(v - v(x, y))/\beta]}\right\}^{-1},$$

where β is a scale parameter. This model is again Fechnerian, whereas one could think of generalizations where the scale parameter becomes choice set dependent, in which case the model is no longer Fechnerian but still symmetric.

We now treat an arbitrary symmetric RUM-CF as the real data-generating process and apply our preference revelation method to the resulting data of choices and response times. It is trivial that these data are rationalizable within the class of symmetric models and hence within the class of all models. More surprising is the fact that our sufficient condition from theorem 1 always recovers the correct preferences.

PROPOSITION 4. Consider an SCF-RT (p, f) that is generated by a symmetric RUM-CF $(u, \tilde{v}, r)$. Then, for any $(x, y) \in D$, $u(x) \geq u(y)$ implies $F(y, x) \text{ } q\text{-FSD } F(x, y)$ and $u(x) > u(y)$ implies $F(y, x) \text{ } q\text{-SFSD } F(x, y)$ for $q = p(x, y)/p(y, x)$.

If we are given a data set that is generated by one of the RUMs commonly employed in the literature, augmented by a chronometric function, then our cautious revealed preference criterion always recovers the correct preferences, despite not using information about the specific data-generating process. Even the analyst who believes in the probit or

logit distribution can work with our criterion. It will yield the same revealed preferences as an application of the full-fledged model if his belief is correct but avoids a mistake if his belief is incorrect.

We add three remarks on this result. First, symmetry of the RUM-CF does not guarantee that the stronger condition in corollary 1 is also satisfied by the SCF-RT.¹⁶ Second, symmetry of the RUM-CF is not necessary for our condition in theorem 1 to be applicable. As the proof of proposition 4 reveals, there are also asymmetric models that generate data from which our criterion recovers preferences correctly. Third, and most importantly, since proposition 4 shows that any data set generated by a symmetric RUM-CF must fulfill the condition in theorem 1, the latter becomes an empirical test of the former. That is, if the condition in theorem 1 is frequently violated in a real data set, this suggests that symmetric RUM-CFs are not an appropriate model for the process generating that data set.

We can go one step further and assume that there is a second source of noise in response times on top of the noise already generated by random utility. The noise could be part of the behavioral model, for example, due to a stochastic chronometric function or randomness in the physiological process implementing the response, or it could be due to imperfect observation by the analyst. We assume that this additional noise is independent from the randomness in utility, so that it does not systematically reverse the chronometric relationship. A common approach in the empirical literature (e.g., Chabris et al. 2009; Fischbacher, Hertwig, and Bruhin 2013; Alós-Ferrer and Ritschel 2018) is to add i.i.d. noise to log response times, where taking the log ensures that response times remain nonnegative. An equivalent way of modeling this is by means of multiplicative noise, which is technically convenient for our purpose. Formally, a RUM with a noisy chronometric function (RUM-NCF) can be obtained from a RUM-CF by letting the response time when x is chosen over y become the random variable

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

conditional on $\tilde{v}(x, y) > 0$. Here, $\tilde{\eta}$ is a nonnegative random variable with mean 1 that is assumed to be i.i.d. according to a density h on $\mathbb{R}_+$. The probability of a response time of at most t , conditional on x being chosen over y , is now the probability that the realized utility difference is at least $r^{-1}(t/\tilde{\eta})$, conditional on that difference being positive. Hence, for an SCF-RT (p, f) that is generated by a RUM-NCF $(u, \tilde{v}, r, \tilde{\eta})$, we have

¹⁶ The example of an SCF-RT presented in app. C, which violates this stronger condition, is actually generated by a symmetric RUM-CF. This RUM-CF features a bimodal utility difference distribution. It is possible to show that symmetry and unimodality together imply that the stronger sufficient condition in corollary 1 is always satisfied. Both the probit and the logit model are unimodal, so working with the stronger condition comes with no loss if either of these models is the data-generating process.

$$F(x, y)(t) = \frac{\int_0^\infty [1 - G(x, y)(r^{-1}(t/\eta))]h(\eta)d\eta}{1 - G(x, y)(0)}$$

for all $t > 0$ and all $(x, y) \in D$, which is analogous to equation (1).

Our preference revelation approach based on RUM-CFs is misspecified when the real data-generating process is a RUM-NCF, because the additional noise is erroneously explained by additional randomness in utility. However, as the next proposition will show, this misspecification is often inconsequential. For the entire class of SCF-RTs that are generated by (and hence are rationalizable in) the class of symmetric RUM-NCFs with full support utility distributions (like probit or logit) and arbitrary response time noise distributions, our previous condition remains the correct criterion for preference revelation.¹⁷

PROPOSITION 5. Consider an SCF-RT (p, f) that is generated by a symmetric RUM-NCF $(u, \tilde{v}, r, \tilde{\eta})$, where each $\tilde{v}(x, y)$ has full support. Then, for any $(x, y) \in D$, $u(x) \geq u(y)$ implies $F(y, x) \text{ } q\text{-FSD } F(x, y)$ and $u(x) > u(y)$ implies $F(y, x) \text{ } q\text{-SFSD } F(x, y)$ for $q = p(x, y)/p(y, x)$.

The proof rests on the insight that q -FSD is invariant to independent perturbations. Whenever an SCF-RT (p, f) satisfies $F(y, x) \text{ } q\text{-FSD } F(x, y)$, then the SCF-RT $(p, \hat{f})$ obtained after perturbing response times by log-additive or multiplicative noise still satisfies $\hat{F}(y, x) \text{ } q\text{-FSD } \hat{F}(x, y)$. The case of RUM-NCFs obtained from symmetric RUM-CFs is a natural application of this insight. However, the robustness of our preference revelation criterion holds more generally for perturbations of any data-generating process for which the criterion has bite. This could be a RUM that is not symmetric, or it could be one of the sequential sampling models studied in the next section.

B. The View from Psychology

A different way of generating stochastic choices and response times is by means of a sequential sampling model, as used extensively in psychology and neuroscience. The basic building block for binary choice problems is the DDM of Ratcliff (1978). A DDM with constant boundaries is given by a drift rate $\mu \in \mathbb{R}$, a diffusion coefficient $\sigma^2 > 0$, and symmetric barriers B and $-B$, with $B > 0$. A stochastic process starts at $Z(0) = 0$ and evolves over time according to a Brownian motion

¹⁷ Without full support of the utility difference distributions, some response times may arise only because of the additional noise but could never be generated by a realized utility difference. The distribution of those response times does not obey the chronometric relationship and would be uninformative of utility.

$$dZ(t) = \mu dt + \sigma dW(t).$$

The process leads to a choice of x (y) if the upper (lower) barrier is hit first, the response time being the time at which this event occurs.

Although the DDM, as a model anchored in psychology, usually does not make reference to underlying utilities, $Z(t)$ is often interpreted as the difference in spiking rates between neurons computing values for the competing options. Hence, it is natural to introduce a link to utility by assuming that the drift rate is determined such that $\mu = \mu(x, y) = -\mu(y, x) \geq 0$ if and only if $v(x, y) \geq 0$. This way, the DDM generates stochastic choices and response times from an underlying deterministic utility function $u: X \rightarrow \mathbb{R}$.

The stochastic path of $Z(t)$ can be interpreted as the accumulation of evidence in favor of one or the other option as the brain samples past (episodic) information. Recent research has shown that evidence accumulation models, like the DDM, actually represent optimal decision-making procedures under neurologically founded constraints, but optimality requires that the barriers are not constant but rather collapse towards zero as t grows to infinity (Tajima, Drugowitsch, and Pouget 2016). Similarly, Fudenberg, Strack, and Strzalecki (2018) model optimal sequential sampling when utilities are uncertain and gathering information is costly and find that the range for which the agent continues to sample should decrease as t grows. A partial intuition for this result is that a value of $Z(t)$ close to zero for small t carries little information, while a value of $Z(t)$ close to zero for large t indicates that the true utilities are most likely close to each other, and hence sampling further evidence has little value. To reflect this idea, a *DDM with collapsing boundaries* works in the exact same way as a DDM with constant boundaries, with the only difference being that the barriers are given by a continuous and strictly decreasing function $b: \mathbb{R}_+ \rightarrow \mathbb{R}_{++}$ such that $\lim_{t \rightarrow \infty} b(t) = 0$. That is, x is chosen if $Z(t)$ hits the upper barrier $b(t)$ before hitting the lower barrier $-b(t)$, and y is chosen if the converse happens, with the response time being the first crossing time (see fig. 2).

We now treat a DDM with an underlying utility function as the real data-generating process and again apply our preference revelation method to the resulting data of choices and response times.¹⁸ The following proposition, the proof of which relies on a result by Fudenberg, Strack, and

¹⁸ For the case of constant boundaries, closed-form solutions for choice probabilities and response time distributions generated by the DDM are known (see, e.g., Palmer, Huk, and Shadlen 2005). Closed-form solutions are not available for the case of collapsing boundaries. Webb (2019) explores the link between bounded accumulation models as the DDM and RUMs and shows how to derive distributional assumptions for realized utilities of the latter if the true data-generating process is of the DDM form.

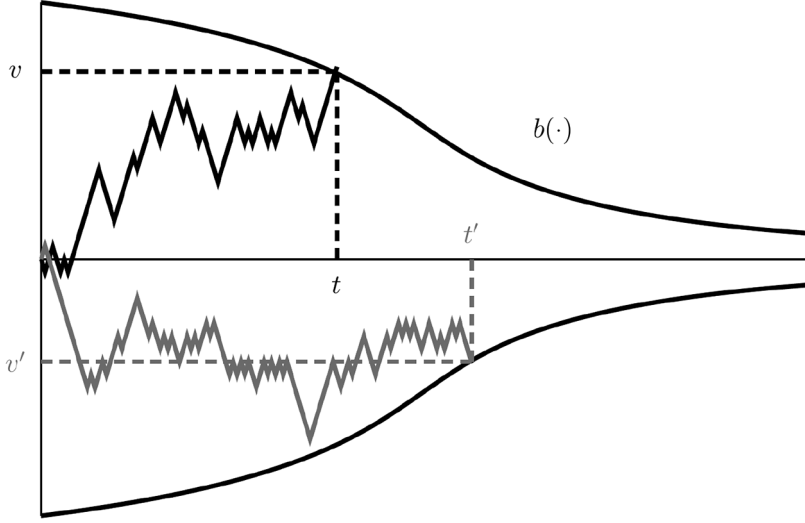


FIG. 2.—Illustration of a DDM with collapsing decision boundaries.

Strzalecki (2018), shows that our sufficient condition from theorem 1 is again tight and always recovers the correct preferences.

PROPOSITION 6. Consider an SCF-RT (p, f) that is generated by a DDM with constant or collapsing boundaries and underlying utility function u . Then, $u(x) \geq u(y)$ implies $F(y, x) \not\prec_q F(x, y)$ and $u(x) > u(y)$ implies $F(y, x) \prec_q F(x, y)$ for $q = p(x, y)/p(y, x)$.

To understand the intuition behind the proof, recall from our discussion of theorem 1 that the condition ensuring a revealed preference can be reformulated as $P(x, y)(t) \geq P(y, x)(t)$ for all t ; that is, the probability of choosing x over y before any prespecified response time t should be larger than the probability of choosing y over x before t . In a DDM, $u(x) > u(y)$ means that the drift rate $\mu(x, y) > 0$ favors x , since the upper barrier reflects a choice of x . Hence, the probability of hitting the upper barrier first before any prespecified response time t is indeed larger than the probability of hitting the lower barrier first. The proof actually establishes that the stronger condition in corollary 1 is always satisfied by an SCF-RT that is generated by a DDM. Furthermore, the result would continue to hold for DDMs with more general boundary functions that are not necessarily constant or collapsing, but these models do not always generate well-behaved choices and have received less attention in the literature.

We remark that, as in the case of proposition 4, proposition 6 is of independent interest. Since it states that any data set generated by a DDM

must fulfill the condition in theorem 1, the latter becomes an empirical test of the former. That is, if the condition in theorem 1 is frequently violated in a real data set, this suggests that DDMs are not an appropriate model for the process generating that data set.

We conclude the section with remarks about rationalizability. An SCF-RT that is generated by a DDM is always rationalizable within the class of all RUM-CFs when we consider just two options, $D = \{(x, y), (y, x)\}$, which is the setting in which DDMs are typically applied. This is a corollary of the following more general result.

PROPOSITION 7. Suppose that $D = \{(x, y), (y, x)\}$. Then, any SCF-RT (p, f) is rationalizable within the class of all RUM-CFs.

In the proof, we fix an arbitrary chronometric function r that satisfies definition 5 (and has $r(v) = 0$ for large v) and construct an associated density $g(x, y)$ such that the data are rationalized. The construction is particularly illuminating for an SCF-RT that is generated by a DDM with collapsing boundaries. In that case, it is very natural to choose the chronometric function of the RUM-CF as the inverse of the boundary function of the DDM. The associated density $g(x, y)$ then describes the distribution of the value of $Z(t)$ at the endogenous decision point (see again fig. 2). This interpretation is in line with Fudenberg, Strack, and Strzalecki (2018), who argue that $Z(t)$ is a signal about the true utility difference. In other words, if we think of the chronometric function as the inverse of the collapsing boundary function, then realized utility differences can be interpreted as realized signals about the underlying deterministic utility difference.

With more than two options, the question of rationalizability is less straightforward. It is not clear that the above construction yields utility difference distributions that are consistent with a single utility function u . For SCF-RTs generated by DDMs, this problem is further complicated by the fact that there is no agreed upon discipline on how utility differences $v(x, y)$ map into drift rates $\mu(x, y)$, beyond the basic ordinal requirement that $\mu(x, y) \geq 0$ if and only if $v(x, y) \geq 0$. Hence, rationalizability of an arbitrary DDM-generated data set in terms of a RUM-CF is not guaranteed. Our approach, however, can be generalized naturally to solve this problem. Define a *generalized RUM with a chronometric function* (GRUM-CF) by replacing RUM.1 in definition 2 by the weaker requirement

$$\mathbb{E}[\tilde{v}(x, y)] \geq 0 \Leftrightarrow u(x) - u(y) \geq 0,$$

keeping everything else unchanged. That is, realized utility differences (or “decision values,” as in the recent neuroeconomics literature; see, e.g., Glimcher and Fehr 2013, chap. 8–10) are interpreted as signals about ordinal preferences rather than cardinal utility. It is easy to see that theorem 1 remains valid for preference revelation within the class of all GRUM-CFs. Furthermore, any SCF-RT generated by a DDM with constant

or collapsing boundaries and underlying utility function u is rationalizable in the class of all GRUM-CFs.¹⁹

V. Empirical Application

In this section, we apply our results to an experiment by Clithero (2018), whom we thank for sharing the data. In the first phase of this experiment, 31 subjects made repeated binary choices between various snack food items and one fixed reference item. Each subject made decisions for 17 such items, and each decision was repeated 10 times. Response times were recorded. In a second phase of the experiment, each subject was confronted once with each of the possible binary choices between the 17 nonreference items from the first phase, resulting in $17 \times 16/2 = 136$ choice problems per subject.

Our strategy of analysis is as follows. First, we use the choice and response time data from the first phase of the experiment to illustrate theorem 1. We show that the q -FSD condition is fulfilled in substantially more than half of the choice problems. That is, theorem 1 has significant empirical bite, allowing for preference identification in the absence of any assumptions on utility noise. At the same time, there remains a nonnegligible fraction of choice problems in which q -FSD is violated, indicating that these choices are not typical realizations of RUM-CFs with symmetric noise distributions or of DDMs. Second, we take full advantage of the experiment's design, where choices in the first phase are always made against the same reference item, and use theorem 2 to derive a complete preference ordering for each subject, relying only on the choice and response time data from the first phase. We then use these preferences to predict choices in the second phase of the experiment. That is, we check the accuracy of the out-of-sample predictions embodied in theorem 2, which rest on no distributional assumptions other than symmetry. The prediction accuracy is remarkable. It clearly outperforms the accuracy of conventional logit models, and it is comparable to computation-intensive DDMs. Third, we conduct an analogous analysis relying on theorem 3 to predict choice probabilities for the second phase of the experiment, assuming a Fechnerian structure of the utility noise. We again achieve a remarkable prediction accuracy.

¹⁹ This approach could be extended even further, building a connection to the economic literature on consumer theory without transitive preferences. Shafer (1974) showed that every complete, continuous, and strongly convex binary relation R (not necessarily transitive) on a Euclidean space can be represented by a continuous, real-valued, two-variable function v such that xRy if and only if $v(x, y) \geq 0$, with $v(x, y) = -v(y, x)$. Hence, one could replace u in the definition of RUM-CFs and DDMs by a complete but not necessarily transitive binary relation r on x . The appropriate reformulation of RUM.1 would be $\mathbb{E}[\tilde{v}(x, y)] \geq 0 \Leftrightarrow xRy$, and DDMs could be linked to R by $\mu(x, y) \geq 0 \Leftrightarrow xRy$.

The purpose of this section is twofold. On the one hand, we illustrate how our methods can be applied in practice. On the other hand, we demonstrate that our theoretical results are empirically relevant, reveal the limits of assumptions made in conventional models, and outperform conventional models that do not use response times in out-of-sample prediction.

A. Applying Theorem 1

We start with an application of theorem 1 to the first phase of the experiment. Our result says that a revealed preference for x over y can be deduced if $F(y, x) \text{ } q\text{-FSD } F(x, y)$ for $q = p(x, y)/p(y, x)$, which can be rewritten as

$$p(x, y)F(x, y)(t) - p(y, x)F(y, x)(t) \geq 0 \text{ for all } t \geq 0.$$

This statement makes use of exact choice probabilities and the full distributions of response times. Since our data set is finite, we need to translate the statement into its empirical counterpart. To do so, we replace the choice probabilities by empirical choice frequencies and replace the response time distributions by nonparametric kernel density estimates.²⁰

Fix a subject and a snack food item z and call the reference item r . Suppose that choice was actually stochastic, so the subject did not make the same choice in all 10 repetitions of the decision problem. The data set contains 77 such decision problems involving 770 choices. We then plot the function

$$\hat{D}(t) = \hat{p}(z, r)\hat{F}(z, r)(t) - \hat{p}(r, z)\hat{F}(r, z)(t),$$

where $\hat{p}(z, r)$ and $\hat{p}(r, z)$ are the empirical choice frequencies across the 10 repetitions and $\hat{F}(z, r)$ and $\hat{F}(r, z)$ are CDFs obtained from nonparametric kernel density estimates of the associated response times.²¹ If $\hat{D}(t) \geq 0$ for all $t \geq 0$, we can deduce a revealed preference for item z over the reference r . Similarly, if $\hat{D}(t) \leq 0$ for all $t \geq 0$, we deduce a revealed

²⁰ An alternative approach is to replace the response time distributions by the empirical CDFs, i.e., step functions with steps at the observed response times. This alternative approach yields results similar to the ones based on kernel estimates, and we report them in the footnotes.

²¹ The kernel density estimates were performed in Stata using the `akdensity` module, which delivers CDFs as output. We estimate the distribution of log response times to avoid boundary problems arising from $t \geq 0$. The estimates use an Epanechnikov kernel with optimally chosen nonadaptive bandwidth. For the case where some choice is made only one out of 10 times—and hence 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.

preference for r over z . If $\hat{D}(t)$ intersects the X -axis at some t , a revealed preference cannot be deduced without distributional assumptions.

Figure 3 shows the function $\hat{D}(t)$ for all stochastic choices of four selected subjects in the data set, normalized so that $\hat{D}(t) \geq 0$ for large t and plotting response times on a logarithmic scale.

There is, of course, a large heterogeneity among subjects. For instance, the subject in the top left panel exhibited stochastic choice for only two of the 17 snack food items. Our q -FSD condition is fulfilled in one of these cases (blue line) but not in the other (red line). The subject in the top right panel exhibited stochastic choice for 13 items, with q -FSD being satisfied in 10 of the cases. Hence, we obtain a distribution-free revealed preference for more than three-quarters of the stochastic choices of this subject.

The subject in the bottom left panels displays stochastic choice for three snacks, of which two fulfill our revealed preference condition. The third case is interesting and represents a case where q -FSD is clearly violated. The blue line initially falls below zero because the quickest choices favor one of the options (the reference r in that case), but it eventually increases above zero because slower choices favor the other option (the snack item z). Overall, the snack z is chosen in six out of 10 cases, but the countervailing pattern of response times makes it impossible to conclude that the distribution of the noise is symmetric enough to obtain a revealed preference. It is worth remarking again that patterns like this one are precluded, except as noisy realizations in a finite data set, both by RUM-CFs with symmetric distributions (as shown by our proposition 4) and by DDMs (as shown by our proposition 6).

The subject in the bottom right panel is particularly interesting. Of the four snacks for which choice was stochastic, three violate q -FSD. The fourth one, however, is such that both the item z and the reference r were chosen exactly five times; hence, the blue line becomes zero for large t . Yet it lies weakly above zero for all t , with many inequalities being strict, meaning that a strict preference (for the reference r in that case) is revealed according to our criterion, while an approach based on the symmetry assumption would conclude an indifference (see our discussion in n. 12).

Across all subjects and decision problems in which choice was stochastic, our q -FSD condition is fulfilled in 61.0% of the cases.²² Put differently, in substantially more than half of the stochastic choice problems, a preference is revealed without distributional assumptions. We take this as evidence of the empirical relevance of the q -FSD criterion. We also note that

²² With CDFs estimated by their empirical step function counterparts, this number is 64.9%. If we do not count as a violation the cases where $\hat{D}(t)$ intersects the X -axis right after the quickest choice—which can happen by chance whenever the quickest choice is in the wrong direction—the number increases to 87.0%.

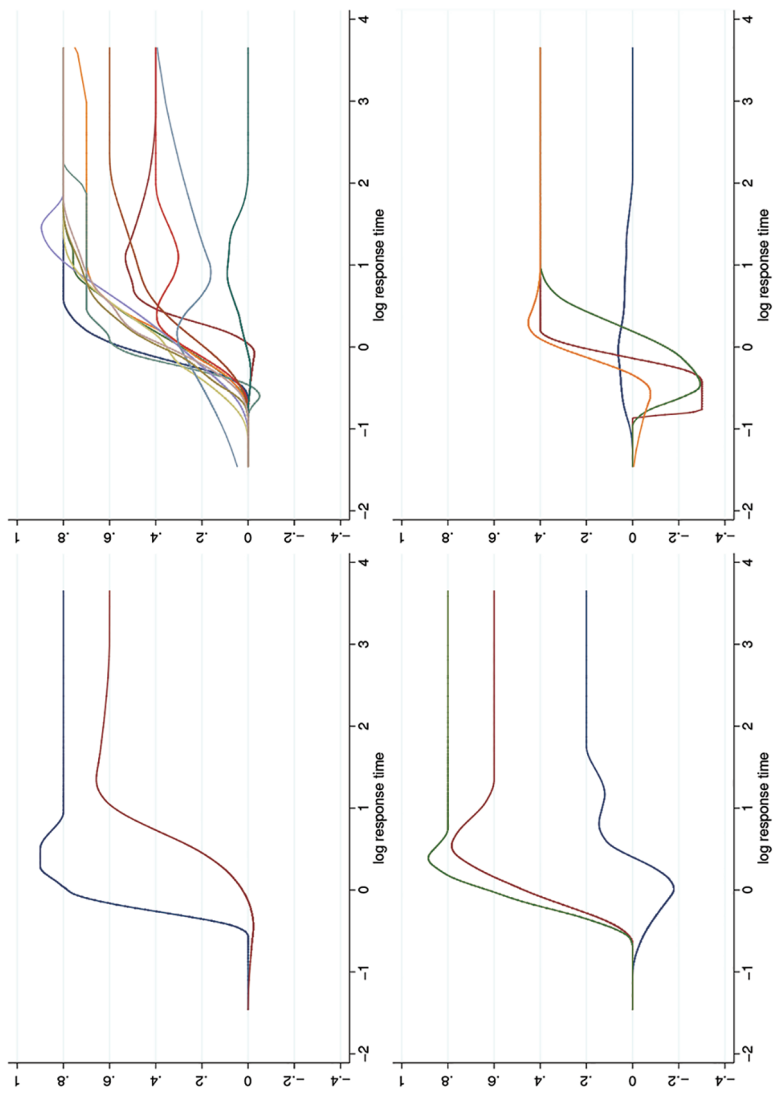


FIG. 3.—Functions $\hat{D}(t)$ for four selected subjects.

classical first-order stochastic dominance obtains in only 35.1% of the cases, and in more than half of these cases (15 out of 27) the less frequently chosen option is actually the faster one, so that there is no systematic relationship between choice frequencies and classical first-order stochastic dominance. Finally, the 39.0% of the cases where q -FSD is violated suggests that neither symmetric RUM-CFs nor DDMs are appropriate models for the data-generating process.

B. Applying Theorem 2

Theorem 2 allows us to make out-of-sample predictions. The experiment of Clithero (2018) is particularly well suited for this purpose, since all first-phase choices were made between different snack food items and a fixed reference item, while in the second phase each possible pair of nonreference items was presented once as a binary decision problem. In a setting like this, we can use the first phase to derive complete preferences and use them to predict choices in the second phase, making no distributional assumptions other than symmetry.

The key to the application is realizing that theorem 2 (plus transitivity) reveals the entire preference relation if one has choice and response time data for the choice between each item z and one fixed reference item r . By proposition 2, all items z with $p(z, r) > 1/2$ are revealed to be strictly better than the reference r . Theorem 2, in turn, implies that items better than r are ordered inversely by $\theta(z, r)$. That is, for any z and z' with $p(z, r) > 1/2$ and $p(z', r) > 1/2$, z is revealed to be strictly preferred over z' if $\theta(z, r) < \theta(z', r)$. Similarly, all items z with $p(z, r) < 1/2$ are strictly worse than r and are ordered by $\theta(r, z)$. Last, any item z with $p(z, r) = 1/2$ is revealed to be indifferent to r .

To deduce the entire preference relation from first-phase choices, we need to identify the empirical counterparts $\hat{\theta}(z, r)$ of the percentiles $\theta(z, r)$ and analogously for $\theta(r, z)$. Again, we rely on the nonparametric kernel density estimates of the response time distributions and define $\hat{\theta}(z, r)$ such that

$$\hat{F}(z, r)(\hat{\theta}(z, r)) = \frac{0.5}{\hat{p}(z, r)}$$

and analogously for $\hat{\theta}(r, z)$.²³ This empirical percentile is also well defined if choice was deterministic, in which case it becomes the median response time of the option that was chosen in all 10 repetitions. While our

²³ More precisely, we estimate $\hat{F}(z, r)(t)$ at all response times t that were ever observed in the entire data set. Then we define $\hat{\theta}(z, r)$ as the average between the largest of the t for which $\hat{F}(z, r)(t) \leq 0.5/\hat{p}(z, r)$ and the smallest of the t for which $\hat{F}(z, r)(t) \geq 0.5/\hat{p}(z, r)$. With the alternative approach based on empirical step function CDFs, the condition $\hat{F}(z, r)(t) = 0.5/\hat{p}(z, r)$ is always satisfied for an entire interval of t , corresponding to one step of the CDF. We then define $\hat{\theta}(z, r)$ as the midpoint of this interval.

theoretical analysis does not cover the case of truly deterministic choice, it predicts that the percentile converges to the median as $p(z, r) \rightarrow 1$. We implement this prediction by ranking also the deterministically chosen options based on $\hat{\theta}(z, r)$ or $\hat{\theta}(r, z)$.

We use the so-derived revealed preference relations to predict choices in the second phase, that is, strictly out of sample. With symmetric utility noise, z is chosen over z' with a probability larger than $1/2$ if and only if z is preferred over z' . Hence, we code a choice of z over z' in the second phase as correctly predicted if z was revealed to be preferred over z' in the first phase.²⁴ Despite the fact that second-phase choices are noisy themselves and the data set contains only one decision for each pair of items, the accuracy of our prediction is remarkable. Across all subjects and binary decision problems, our predictions are correct in 80.7% of the cases.²⁵

Clithero (2018) conducted the same prediction exercise on the basis of additional structural assumptions. First, he fitted a conventional logit model to the first-phase choices and used the individual parameters to predict second-phase choices.²⁶ This method yielded a prediction accuracy of only 73.8%. This accuracy of the logit model is significantly lower than for our approach based on theorem 2 (paired t -test based on subject-level prediction accuracy that we obtained from Clithero, $t = 5.4046$, $p < .0001$). Second, Clithero (2018) fitted a fully parametric DDM, using first-phase choices and response times. This method yielded a prediction accuracy of 81.2%, which is statistically indistinguishable from our approach (paired t -test, $t = -0.7781$, $p = .4426$). Hence, we achieve the same accuracy as a DDM but without any parametric estimation and with no structural assumptions beyond symmetry of noise and the chronometric relationship.

Our prediction accuracy is even higher for a subset of decisions. The second phase of the experiment collected only one decision for each pair of snacks. Since choices are stochastic, prediction accuracy must be limited simply because of the inherent noise. Previous research has documented that noisier decisions, which are thus more difficult to predict accurately, have longer response times on average (e.g., Chabris et al. 2009; Alós-Ferrer and Garagnani 2018). We do have the response times of the second-phase decisions, which we did not use so far. As a simple exercise, we divided the data in four response time quartiles for each subject and reexamined the corresponding subsets of decisions. For the slowest decisions, prediction accuracy drops to 68.5%. For the second-slowest quartile, it rises to

²⁴ The revealed preference relations that we obtain are always strict, except for one subject and one pair of items. We ignore this single case of indifference in our prediction evaluation.

²⁵ With empirical step function CDFs, the prediction is correct in 80.0% of the cases.

²⁶ The fitting approach relies on a Markov chain Monte Carlo method assuming comparability of choices across individuals. Comparability across individuals is necessary because, if not, $\hat{p}(z, r) \in \{0, 1\}$ would deliver no information. The Monte Carlo approach generates information from the fact that $\hat{p}(z, r) \notin \{0, 1\}$ for some other individual.

79.0%. For the second-fastest quartile, it is already at 86.7%. Finally, for the quartile containing the fastest decisions, prediction accuracy is at a remarkable 88.7%. This shows the actual potential of our method.

C. Applying Theorem 3

If one is willing to assume that utility noise has a Fechnerian structure (but still not commit to a specific form, such as logit or probit), theorem 3 allows an even more fine-tuned out-of-sample prediction in the form of choice probabilities $\bar{p}(x, y)$ for unobserved choices. We apply this result to the same data set as above.

The procedure is analogous to that illustrated in section V.B. For each pair of snacks (z, z') , we obtain a probability prediction from the formula in theorem 3 by replacing all expressions by their empirically estimated counterparts, just like before. However, we can use theorem 3 to independently predict either $\bar{p}(z, z')$ or $\bar{p}(z', z)$, and the two predictions are not guaranteed to satisfy $\bar{p}(z, z') = 1 - \bar{p}(z', z)$. Put differently, we may obtain two different predictions for the same choice probability by swapping the roles of x and y in theorem 3. The issue does not arise with a large and rationalizable data set like in our theoretical analysis, but it will arise in finite noisy data. It turns out that differences are small in our data set. On average across all second-phase decisions, the difference between $\bar{p}(z, z')$ and $1 - \bar{p}(z', z)$ is 0.04 and thus less than 5 percentage points.²⁷ For the following evaluation, we take the average of the two predictions.

To measure the accuracy of our prediction, we compute the mean absolute error across all decisions in the second phase of the experiment. If z was chosen over z' , the absolute error of the prediction in this decision is defined as $1 - \bar{p}(z, z')$, which implies a smaller error if the observed choice was predicted to be more likely. If z' was chosen over z , the absolute error is $\bar{p}(z, z')$. Proceeding in this way, we obtain a mean absolute error of 0.237.²⁸

The logit model of Clithero (2018) yielded a mean absolute error of 0.263. Compared with our method, this error is significantly larger (paired t -test, $t = -3.7208$, $p = .0008$), so that our response-time based approach again outperforms the conventional RUM. The DDM of Clithero (2018) achieved a mean absolute error of 0.209, which is significantly lower than for our method (paired t -test, $t = 7.3948$, $p < .0001$). We emphasize that our nonparametric method yields its high accuracy in a computationally straightforward way and does not require structural estimation like a DDM.

Our prediction is again more accurate for faster decisions. We predict probabilities closer to 0 or 1 for second-phase decisions that turn out to

²⁷ With empirical step function CDFs, the average difference is 0.06. Since probability predictions are on a much coarser grid in that case, there is actually no difference at all in 57.2% of the cases.

²⁸ With empirical step function CDFs, the number is 0.235.

happen faster, in line with the literature cited earlier.²⁹ For the slowest decisions, the mean absolute prediction error is 0.354. For the second-slowest quartile, it falls to 0.253. For the second-fastest quartile, it is already at 0.188. Finally, for the quartile containing the fastest decisions, the error is at only 0.152.

VI. Relation to the Literature

A. *Stochastic Choice and Response Times: History*

The realization that choice is inherently stochastic has permeated economics for decades and psychology even longer.³⁰ Models of stochastic choice in economics view SCFs—the most direct theoretical counterpart of observable choice frequencies—as a primitive. Block and Marschak (1960) posed the question of when an SCF can be represented as a distribution over preferences, with the interpretation that either choices are generated by randomly sampled preferences at the individual level or they reflect the choices of a population of randomly sampled agents. Not every SCF can be viewed as such a distribution, but Falmagne (1978), Barberá and Pattanaik (1986), and McFadden and Richter (1990) derived necessary and sufficient conditions under which this is possible. Those random preference models are behaviorally equivalent to RUMs, which assume the existence of a utility function that is perturbed by an error term. For finite sets of alternatives, Block and Marschak (1960) showed that choice probabilities can be rationalized by a random preference model if and only if they can be rationalized by a RUM. The applied literature overwhelmingly relies on the second approach, because it allows for parametric estimation once one fixes functional forms for the utility function and the noise term. RUMs have a long history, with contributions including Marschak (1960) and going back to Thurstone (1927), but they first became widespread in economics after being used in transportation science (Ben-Akiva and Lerman 1985), as McFadden (2001) explains in his Nobel Prize lecture. It was Thurstone (1927) in psychology who first introduced what is known today as the probit model, where utility errors are assumed to be normally distributed. This is the first example of a Fechnerian model, as described in our section III.C. In these models, choice probabilities can be derived from a fixed CDF of utility differences (see also Debreu 1958). A different example is the logit

²⁹ For the slowest quartile of decisions, the average predicted probability of the more likely choice is 0.74. It becomes 0.81 for the second-slowest quartile, 0.85 for the second-fastest quartile, and 0.89 for the fastest quartile.

³⁰ The extent to which this is true seems to be often underestimated. For instance, Samuelson (1938), which is widely considered to be one of the seminal contributions to have started the modern revealed preference approach, was, as its title indicated, a note on the earlier paper of Georgescu-Roegen (1936), which explicitly viewed consumer choice as stochastic.

model, where one assumes a logistic distribution for utility differences. The logit model is particularly interesting because choice probabilities can be written in a closed form that is equivalent to the celebrated choice model of Luce (1959), where the choice probability of an option is proportional to a normalized value that can be interpreted as the utility of the option (for a detailed treatment, see Anderson, Thisse, and De Palma 1992, chap. 1). Generating choice probabilities from decision values by assuming a logit function (or softmax) is also a common approach in decision neuroscience.

Parallel to developments in economics and related sciences, cognitive psychology—inspired by the same earlier observations and models—followed a different path. The motivation was to explain both choice frequencies and data associated with the choice process, such as response times. A key development in this literature was the DDM of Ratcliff (1978) described in section IV.B (see also Ratcliff and Rouder 1998). The DDM predicts logit probabilities in binary choice and, as argued by Woodford (2014), could be seen as providing a foundation for the latter. The model essentially evolved from a chain of contributions adapting the sequential ratio probability test of Wald (1945) as a descriptive model of choice and has grown to become a veritable standard in cognitive psychology and neuroscience. Although the reasons for its original success are probably linked to its flexibility in fitting data, it has been argued to reflect neural processes implementing choices (Shadlen and Kiani 2013; Shadlen and Shohamy 2016). Recent developments have integrated attentional processes in the form of eye-tracking data into the DDM (Krajbich, Armel, and Rangel 2010; Krajbich and Rangel 2011). Empirical estimates of these attentional DDMs indicate that nonattended options tend to be discounted in the process of evidence accumulation. As discussed in the text, models such as the one in Ratcliff (1978) can also be interpreted as models where latent utility differences are discovered by the decision maker, and that discovery takes time (Fudenberg, Strack, and Strzalecki 2018). It is precisely this reinterpretation that provides a powerful intuition for the chronometric function: it takes time to tell apart two close values, while differentiating two values that are far apart is quickly done.

Our work constitutes a bridge between the disparate branches of the literature by providing a framework where response times can be understood as an integral part of RUMs.

B. Response Times: Recent Developments

Our work is related to a recent strand of the literature that makes the empirical point that response time data can help with structural estimation of preferences by using the chronometric function. Konovalov and Krajbich (2019) and Schotter and Trevino (2021) propose to estimate indifference points by using the longest response times in a data set and then

deduce ordinal preference relations from the indifference points. Other studies have shown how response times are indicative of effort allocation (Moffatt 2005) or can be used to improve out-of-sample predictions of choices (Clithero 2018). All those works, however, consider fully specified structural models and add response times to improve the estimation, an approach that can be useful when choice data are scarce or not reliable. In contrast, our paper solves a different problem. As proposition 1 shows, in the absence of unverifiable assumptions about the structure of utility noise, it is impossible to uncover preferences from choices. We show that this fundamental problem can be overcome with response time data. We provide, in particular, a simple and intuitive condition that ensures that preferences can be recovered nonparametrically in the absence of any assumptions on the distribution of utility noise, that is, under conditions where the recovery of preferences fails even with rich choice data.

Echenique and Saito (2017) provide an axiomatization of the chronometric relationship, viewed as a mapping from utility differences to response times, as in our model. They consider deterministic choices and deterministic response times only. Their main interest is a characterization of finite and incomplete data sets that can be rationalized by a deterministic utility function together with a chronometric function. That is, they do not consider stochastic choice or the problems that arise when utility is noisy.

While response times are generally receiving increased attention as a tool to improve economic analysis, detailed studies are still scarce (a review and discussion can be found in Spiliopoulos and Ortmann 2018). Examples include the studies of risky decision-making by Wilcox (1993, 1994), the web-based studies of Rubinstein (2007, 2013), and the study of belief updating by Achtziger and Alós-Ferrer (2014). The study of Alós-Ferrer et al. (2016) uses the chronometric relationship to understand the preference reversal phenomenon (Grether and Plott 1979), where decision makers typically make lottery choices that contradict their elicited certainty equivalents when one of the lotteries has a salient, large outcome. Alós-Ferrer et al. (2016) show that if reversals are due to a bias in the elicitation process rather than in the choice process, choices associated with reversals should take longer than comparable nonreversals. The reason is simply that reversals (where noisy valuations flip) are more likely when the actual utilities are close, and hence, by the chronometric relationship, response times must be longer. The prediction is readily found in the data, providing insights into the origin and nature of reversals.

C. Stochastic Choice: Recent Developments

Our work is related to the recent literature on stochastic choice theory using extended data sets. Caplin and Dean (2015) and Caplin and Martin

(2015) consider data sets that specify choice frequencies as functions of observable signals about an underlying state. Their results focus on rationalizability when the decision maker has a prior and updates it through Bayes' rule after receiving the signals, but state-dependent choice data adds an additional dimension that could potentially help with preference revelation too. Extended data also play a role in the model of Eliaz and Rubinstein (2014), where decision makers observe and can imitate other agents but can additionally observe correlates of their decisions and, in particular, whether those decisions were hasty or not.

Our paper also contributes to the theory of behavioral welfare economics, which aims at eliciting preferences from inconsistent choice data (among others, see Bernheim and Rangel 2009; Masatlioglu, Nakajima, and Ozbay 2012; Rubinstein and Salant 2012; Benkert and Netzer 2018). Most of this literature considers deterministic choices. Two exceptions are Manzini and Mariotti (2014) and Apesteguía and Ballester (2015). Manzini and Mariotti (2014) show that underlying preferences can be identified when stochastic choice is due to stochastic consideration sets. Apesteguía and Ballester (2015) propose as a welfare measure the preference relation that is closest (in a certain, well-defined sense) to the observed stochastic choices. Similar to our results in section IV, they show that this procedure recovers the true underlying preference if the data are generated by RUMs fulfilling a monotonicity condition. To the best of our knowledge, this literature has not yet discovered the value of response time data for preference revelation.

The difficulty of a choice problem can be influenced by additional factors on top of the utility difference between the options. For instance, if the options are multidimensional, then a choice problem involving a dominant alternative may be very simple, generating accurate and quick responses even if the underlying utility difference is small (see, e.g., He and Natenzon 2018). This is a well-known problem for conventional RUMs, because if the utility difference is small, error rates are predicted to be high. However, our definition of RUMs sidesteps this problem. The reason is that the added generality in definition 2 allows for pair-specific random utility differences $\tilde{v}(x, y)$ that are not necessarily simple differences $\tilde{u}(x) - \tilde{u}(y)$ between option-specific random variables. For instance, He and Natenzon (2018) consider moderate utility models relating choice probabilities $p(x, y)$ to utility differences $u(x) - u(y)$ and an additional distance $d(x, y)$, which reflects factors beyond utility differences that may influence the pair-specific choice difficulty. The assumption is that $p(w, x) \geq p(y, z)$ if and only if $(u(w) - u(x))/d(w, x) \geq (u(y) - u(z))/d(y, z)$, which retains the basic regularities of conventional models while allowing for violations of weak stochastic transitivity. Such considerations can be reflected in our setting by working with a formulation like $\tilde{v}(x, y) = u(x) - u(y) + d(x, y)(\tilde{\epsilon}(x) - \tilde{\epsilon}(y))$. A similar point applies if decisions are subject

to frames (e.g., Salant and Rubinstein 2008; Benkert and Netzer 2018) or to pair-specific impulsive tendencies or intuitive processes (Achtziger and Alós-Ferrer 2014), which might influence response times independently of the underlying utilities.

VII. Conclusion

Choice theory has traditionally focused on choice outcomes and has ignored auxiliary data, such as response times. This neglect comes at a cost even for traditional choice-theoretic questions. In the context of stochastic choice, ignoring response time data means discarding information about the distribution of random utility, which then has to be compensated by making distributional assumptions. In this paper, we have developed a suite of tools that utilize response time data for a recovery of preferences without or with fewer distributional assumptions. Table 1 summarizes our main results on what can be learned from a data set. The table illustrates that using response times is always a substitute for making the next stronger distributional assumption. We have also shown that our tools are easy to apply and generate remarkably accurate results.

Throughout most of the paper, we have interpreted SCF-RTs as describing the choices of a single individual who is confronted with the same set of options repeatedly. Random utility then reflects fluctuating tastes or noisy perception of the options. However, our tools also work when the data are generated by a heterogeneous population of individuals, each of whom makes a deterministic choice at a deterministic response time (as in Echenique and Saito 2017). Random utility then reflects a distribution of deterministic utility functions within the population, and response times vary because the difficulty of the choice problem varies with the subjective utility difference. At first glance, a one-to-one translation of our results seems to require the assumption that the same chronometric function applies to all individuals.³¹ However, what we called a noisy chronometric function in section IV can readily be interpreted as a distribution of chronometric functions within the population. In this population interpretation, a revealed preference for x over y means that utilitarian welfare with x is larger than with y . The use of response times is a novel way to approach the long-standing problem of how to measure the cardinal properties of utility that utilitarianism relies on.³²

³¹ Empirical results by Chabris et al. (2009) and Konovalov and Krajbich (2019) indicate that response times (even as little as one observation per individual) can indeed be used to track down parametric differences in utilities across individuals.

³² See d'Aspremont and Gevers (2002) for a discussion. The requirement that each individual's chronometric function is drawn independently from a restricted class (as described in sec. IV) mirrors the interpersonal comparability of utility units that utilitarianism requires.

TABLE 1
SUMMARY OF MAIN RESULTS

Assumptions	Choice Data Only	Choice and Response Time Data
None	Cannot learn anything about preferences (proposition 1)	Learn preferences within sample (theorem 1)
Symmetric noise	Learn preferences within sample (proposition 2)	Also learn preferences for option pairs outside data set (theorem 2)
Fechnerian noise	Learn preferences for option pairs outside data set (proposition 3)	Also learn choice probabilities for option pairs outside data set (theorem 3)

There is a range of interesting questions that we leave for future research. First, our results lend themselves to empirical application. A first natural step was to work with experimental choice data from the lab, where response times are easy to measure. Future work could study real-world data, for example, from online marketplaces, where the time a consumer spends contemplating the options could be (and presumably is already) recorded. A challenge will be to differentiate response time in our sense from other concepts, such as the time required to read information or to deliberate on the consequences of an action, which may have other qualitative predictions (as in Rubinstein 2007, 2013).

Second, we have not attempted a full characterization of rationalizability for arbitrary SCF-RTs beyond those studied in section IV. For the case without response times, characterizations are relatively simple and have been given in the literature.³³ The problem is substantially more involved when response time distributions have to be rationalized too. However, some useful necessary conditions are easy to obtain. For instance, an SCF-RT can be rationalizable in the class of all RUM-CFs only if the preferences revealed according to theorem 1 have no cycles, as otherwise there cannot exist a utility function that is consistent with the revealed preferences. Analogous conditions hold for rationalizability in the classes of symmetric or Fechnerian RUM-CFs. These simple conditions provide a specific test of our response time-based model and allow it to be falsified by the data.

Finally, response times are a particularly simple measure with a well-established relation to underlying preferences, but they may not be the only auxiliary data with that property. Physiological measures—such as pupil dilation, blood pressure, or brain activation—may also carry systematic information about preferences. It is worth exploring to what extent these measures can improve the classical revealed preference approach and should therefore be added to the economics toolbox.

³³ For instance, it can be shown that an SCF is rationalizable in our class of symmetric RUMs if and only if the binary relation R defined in app. B has no cycles, in the sense of Suzumura (1976).

References

- Achtziger, A., and C. Alós-Ferrer. 2014. "Fast or Rational? A Response-Times Study of Bayesian Updating." *Management Sci.* 60 (4): 923–38.
- Agranov, M., and P. Ortoleva. 2017. "Stochastic Choice and Preferences for Randomization." *J.P.E.* 125 (1): 40–68.
- Alós-Ferrer, C., and M. Garagnani. 2018. "Strength of Preference and Decisions under Risk." Working paper, Univ. Zurich.
- Alós-Ferrer, C., D.-G. Granić, J. Kern, and A. K. Wagner. 2016. "Preference Reversals: Time and Again." *J. Risk and Uncertainty* 52 (1): 65–97.
- Alós-Ferrer, C., and A. Ritschel. 2018. "The Reinforcement Heuristic in Normal Form Games." *J. Econ. Behavior and Org.* 152:224–34.
- Anderson, S. P., J.-F. Thisse, and A. De Palma. 1992. *Discrete Choice Theory of Product Differentiation*. Cambridge, MA: MIT Press.
- Apesteguía, J., and M. A. Ballester. 2015. "A Measure of Rationality and Welfare." *J.P.E.* 123:1278–310.
- . 2018. "Monotone Stochastic Choice Models: The Case of Risk and Time Preferences." *J.P.E.* 126 (1): 74–106.
- Apesteguía, J., M. A. Ballester, and J. Lu. 2017. "Single-Crossing Random Utility Models." *Econometrica* 85 (2): 661–74.
- Arrow, K. J. 1959. "Rational Choice Functions and Orders." *Economica* 26:121–27.
- Baldassi, C., S. Cerreia-Vioglio, F. Maccheroni, and M. Marinacci. 2020. "A Behavioral Characterization of the Drift Diffusion Model and Its Multi-Alternative Extension to Choice under Time Pressure." *Management Sci.* 66 (11): 4921–5484.
- Ballinger, T. P., and N. T. Wilcox. 1997. "Decisions, Error and Heterogeneity." *Econ. J.* 107 (443): 1090–105.
- Barberá, S., and P. K. Pattanaik. 1986. "Falmagne and the Rationalizability of Stochastic Choices in Terms of Random Orderings." *Econometrica* 54 (3): 707–15.
- Ben-Akiva, M. E., and S. R. Lerman. 1985. *Discrete Choice Analysis: Theory and Applications to Travel Demand*. Cambridge, MA: MIT Press.
- Benkert, J.-M., and N. Netzer. 2018. "Informational Requirements of Nudging." *J.P.E.* 126 (6): 2323–55.
- Bernheim, B. D., and A. Rangel. 2009. "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics." *Q.J.E.* 124:51–104.
- Block, H. D., and J. Marschak. 1960. "Random Orderings and Stochastic Theories of Responses." In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, edited by I. Olkin, 97–132. Stanford: Stanford Univ. Press.
- Bond, T. N., and K. Lang. 2019. "The Sad Truth about Happiness Scales." *J.P.E.* 127:1629–40.
- Buschena, D., and D. Zilberman. 2000. "Generalized Expected Utility, Heteroscedastic Error, and Path Dependence in Risky Choice." *J. Risk and Uncertainty* 20:67–88.
- Camerer, C. F. 1989. "Does the Basketball Market Believe in the 'Hot Hand'?" *A.E.R.* 79:1257–61.
- Caplin, A., and M. Dean. 2015. "Revealed Preference, Rational Inattention, and Costly Information Acquisition." *A.E.R.* 105 (7): 2183–203.
- Caplin, A., and D. Martin. 2015. "A Testable Theory of Imperfect Perception." *Econ. J.* 125:184–202.
- Cattell, J. M. 1893. "On Errors of Observation." *American J. Psychology* 5 (3): 285–93.
- . 1902. "The Time of Perception as a Measure of Differences in Intensity." *Philosophische Studien* 19:63–68.

- Chabris, C. F., C. L. Morris, D. Taubinsky, D. Laibson, and J. P. Schuldt. 2009. "The Allocation of Time in Decision-Making." *J. European Econ. Assoc.* 7 (2–3): 628–37.
- Clithero, J. A. 2018. "Improving Out-of-Sample Predictions Using Response Times and a Model of the Decision Process." *J. Econ. Behavior and Org.* 148:344–75.
- Cohen, M., and J.-C. Falmagne. 1990. "Random Utility Representation of Binary Choice Probabilities: A New Class of Necessary Conditions." *J. Mathematical Psychology* 34:88–94.
- Dashiell, J. F. 1937. "Affective Value-Distances as a Determinant of Aesthetic Judgment-Times." *American J. Psychology* 50:57–67.
- d'Aspremont, C., and L. Gevers. 2002. "Social Welfare Functionals and Interpersonal Comparability." In *Handbook of Social Choice and Welfare*, edited by K. J. Arrow, A. K. Sen, and K. Suzumura, 450–541. Amsterdam: Elsevier.
- Debreu, G. 1958. "Stochastic Choice and Cardinal Utility." *Econometrica* 26 (3): 440–44.
- Dehaene, S., E. Dupoux, and J. Mehler. 1990. "Is Numerical Comparison Digital? Analogical and Symbolic Effects in Two-Digit Number Comparison." *J. Experimental Psychology: Human Perception and Performance* 16 (3): 626–41.
- Drugowitsch, J., R. Moreno-Bote, A. K. Churchland, M. N. Shadlen, and A. Pouget. 2012. "The Cost of Accumulating Evidence in Perceptual Decision Making." *J. Neuroscience* 32:3612–28.
- Echenique, F., and K. Saito. 2017. "Response Time and Utility." *J. Econ. Behavior and Org.* 139 (15): 49–59.
- Edwards, W. 1954. "The Theory of Decision Making." *Psychological Bull.* 51 (4): 380–417.
- Eliaz, K., and A. Rubinstein. 2014. "A Model of Boundedly Rational 'Neuro' Agents." *Econ. Theory* 57:515–28.
- Falmagne, J.-C. 1978. "A Representation Theorem for Finite Random Scale Systems." *J. Mathematical Psychology* 18 (1): 52–72.
- Fechner, T. G. 1860. *Elemente der Psychophysik [Elements of Psychophysics]*. Leipzig: Breitkopf & Härtel.
- Fischbacher, U., R. Hertwig, and A. Bruhin. 2013. "How to Model Heterogeneity in Costly Punishment: Insights from Responders' Response Times." *J. Behavioral Decision Making* 26 (5): 462–76.
- Fisher, G. 2017. "An Attentional Drift-Diffusion Model over Binary-Attribute Choice." *Cognition* 168:34–45.
- Fudenberg, D., P. Strack, and T. Strzalecki. 2018. "Speed, Accuracy, and the Optimal Timing of Choices." *A.E.R.* 108 (12): 3651–84.
- Georgescu-Roegen, N. 1936. "The Pure Theory of Consumer's Behaviour." *Q.J.E.* 50 (4): 545–93.
- Glimcher, P. W., and E. Fehr. 2013. *Neuroeconomics: Decision Making and the Brain*. 2nd ed. Amsterdam: Elsevier.
- Grether, D. M., and C. R. Plott. 1979. "Theory of Choice and the Preference Reversal Phenomenon." *A.E.R.* 69 (4): 623–38.
- Gul, F., P. Natenzon, and W. Pesendorfer. 2014. "Random Choice as Behavioral Optimization." *Econometrica* 82 (5): 1873–912.
- Gul, F., and W. Pesendorfer. 2006. "Random Expected Utility." *Econometrica* 74 (1): 121–46.
- Haile, P. A., A. Hortaçsu, and G. Kosenok. 2008. "On the Empirical Content of Quantal Response Equilibrium." *A.E.R.* 98 (1): 180–200.
- He, J., and P. Natenzon. 2018. "Moderate Expected Utility." Working paper, Washington Univ. Saint Louis.

- Hey, J. D., and C. Orme. 1994. "Investigating Generalizations of Expected Utility Theory Using Experimental Data." *Econometrica* 62 (6): 1291–326.
- Horan, S., P. Manzini, and M. Mariotti. 2019. "When Is Coarseness Not a Curse? Comparative Statics of the Coarse Random Utility Model." Discussion paper.
- Houthakker, H. S. 1950. "Revealed Preference and the Utility Function." *Economica* 17 (66): 159–74.
- Kahneman, D. 2003. "Maps of Bounded Rationality: Psychology for Behavioral Economics." *A.E.R.* 93 (5): 1449–75.
- Klein, A. S. 2001. "Measuring, Estimating, and Understanding the Psychometric Function: A Commentary." *Attention, Perception, and Psychophysics* 63 (8): 1421–55.
- Konovalov, A., and I. Krajbich. 2019. "Revealed Strength of Preference: Inference from Response Times." *Judgment and Decision Making* 14 (4): 381–94.
- Krajbich, I., C. Armel, and A. Rangel. 2010. "Visual Fixations and the Computation and Comparison of Value in Simple Choice." *Nature Neuroscience* 13 (10): 1292–98.
- Krajbich, I., B. Bartling, T. Hare, and E. Fehr. 2015. "Rethinking Fast and Slow Based on a Critique of Reaction-Time Reverse Inference." *Nature Communications* 6 (7455): 1–9.
- Krajbich, I., D. Lu, C. Camerer, and A. Rangel. 2012. "The Attentional Drift-Diffusion Model Extends to Simple Purchasing Decisions." *Frontiers Psychology* 3:193.
- Krajbich, I., B. Oud, and E. Fehr. 2014. "Benefits of Neuroeconomic Modeling: New Policy Interventions and Predictors of Preference." *A.E.R.* 104 (5): 501–6.
- Krajbich, I., and A. Rangel. 2011. "Multialternative Drift-Diffusion Model Predicts the Relationship between Visual Fixations and Choice in Value-Based Decisions." *Proc. Nat. Acad. Sci. USA* 108 (33): 13852–57.
- Laming, D. 1985. "Some Principles of Sensory Analysis." *Psychological Rev.* 92 (4): 462–85.
- Luce, R. D. 1959. *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.
- . 1986. *Response Times: Their Role in Inferring Elementary Mental Organization*. New York: Oxford Univ. Press.
- Manski, C. F. 1977. "The Structure of Random Utility Models." *Theory and Decision* 8 (3): 229–54.
- Manzini, P., and M. Mariotti. 2014. "Stochastic Choice and Consideration Sets." *Econometrica* 82 (3): 1153–76.
- Marschak, J. 1960. "Binary Choice Constraints on Random Utility Indicators." In *Stanford Symposium on Mathematical Methods in the Social Sciences*, edited by K. J. Arrow, 312–29. Stanford: Stanford Univ. Press.
- Masatlioglu, Y., D. Nakajima, and E. Y. Ozbay. 2012. "Revealed Attention." *A.E.R.* 102 (5): 2183–205.
- McFadden, D. L. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*, edited by P. Zarembka, 105–42. New York: Academic Press.
- . 2001. "Economic Choices." *A.E.R.* 91 (3): 351–78.
- McFadden, D. L., and M. K. Richter. 1990. "Stochastic Rationality and Revealed Preference." In *Preferences, Uncertainty, and Optimality: Essays in Honor of Leonid Hurwicz*, edited by J. S. Chipman, D. L. McFadden, and M. K. Richter, 163–86. Boulder, CO: Westview.
- Moffatt, P. G. 2005. "Stochastic Choice and the Allocation of Cognitive Effort." *Experimental Econ.* 8 (4): 369–88.
- . 2015. *Experimentics: Econometrics for Experimental Economics*. London: Palgrave Macmillan.

- Mosteller, F., and P. Noguee. 1951. "An Experimental Measurement of Utility." *J.P.E.* 59:371–404.
- Moyer, R. S., and R. H. Bayer. 1976. "Mental Comparison and the Symbolic Distance Effect." *Cognitive Psychology* 8 (2): 228–46.
- Moyer, R. S., and T. K. Landauer. 1967. "Time Required for Judgements of Numerical Inequality." *Nature* 215 (5109): 1519–20.
- Palmer, J., A. C. Huk, and M. N. Shadlen. 2005. "The Effect of Stimulus Strength on the Speed and Accuracy of a Perceptual Decision." *J. Vision* 5:376–404.
- Ratcliff, R. 1978. "A Theory of Memory Retrieval." *Psychological Rev.* 85:59–108.
- Ratcliff, R., and J. N. Rouder. 1998. "Modeling Response Times for Two-Choice Decisions." *Psychological Sci.* 9 (5): 347–56.
- Ratcliff, R., P. L. Smith, S. D. Brown, and G. McKoon. 2016. "Diffusion Decision Model: Current Issues and History." *Trends Cognitive Sci.* 20:260–81.
- Rieskamp, J., J. R. Busemeyer, and B. A. Mellers. 2006. "Extending the Bounds of Rationality: Evidence and Theories of Preferential Choice." *J. Econ. Literature* 44 (3): 631–61.
- Rubinstein, A. 2007. "Instinctive and Cognitive Reasoning: A Study of Response Times." *Econ. J.* 117 (523): 1243–59.
- . 2013. "Response Time and Decision Making: An Experimental Study." *Judgment and Decision Making* 8 (5): 540–51.
- Rubinstein, A., and Y. Salant. 2012. "Eliciting Welfare Preferences from Behavioural Data Sets." *Rev. Econ. Studies* 79:375–87.
- Salant, Y., and A. Rubinstein. 2008. "(A, f): Choice with Frames." *Rev. Econ. Studies* 75:1287–96.
- Samuelson, P. A. 1938. "A Note on the Pure Theory of Consumer's Behavior." *Economica* 56 (17): 61–71.
- Schotter, A., and I. Trevino. 2021. "Is Response Time Predictive of Choice? An Experimental Study of Threshold Strategies." *Experimental Econ.* 24:87–117.
- Shadlen, M. N., and R. Kiani. 2013. "Decision Making as a Window on Cognition." *Neuron* 80:791–806.
- Shadlen, M. N., and D. Shohamy. 2016. "Decision Making and Sequential Sampling from Memory." *Neuron* 90:927–39.
- Shafer, W. J. 1974. "The Nontransitive Consumer." *Econometrica* 42:913–19.
- Spiliopoulos, L., and A. Ortmann. 2018. "The BCD of Response Time Analysis in Experimental Economics." *Experimental Econ.* 21 (2): 383–433.
- Suzumura, K. 1976. "Remarks on the Theory of Collective Choice." *Economica* 43:381–90.
- Tajima, S., J. Drugowitsch, and A. Pouget. 2016. "Optimal Policy for Value-Based Decision-Making." *Nature Communications* 7:12400.
- Thurstone, L. L. 1927. "A Law of Comparative Judgment." *Psychological Rev.* 34:273–86.
- Tsetos, K., R. Moran, J. Moreland, N. Chater, M. Usher, and C. Summerfield. 2016. "Economic Irrationality Is Optimal during Noisy Decision Making." *Proc. Nat. Acad. Sci. USA* 113 (11): 3102–7.
- Tversky, A. 1969. "Intransitivity of Preferences." *Psychological Rev.* 76:31–48.
- Wald, A. 1945. "Sequential Tests of Statistical Hypotheses." *Ann. Mathematical Statist.* 16 (2): 117–86.
- Webb, R. 2019. "The (Neural) Dynamics of Stochastic Choice." *Management Sci.* 64 (1): 230–55.
- Wichmann, A. F., and N. J. Hill. 2001. "The Psychometric Function: I. Fitting, Sampling, and Goodness of Fit." *Attention, Perception, and Psychophysics* 63 (8): 1293–313.

Wilcox, N. T. 1993. "Lottery Choice: Incentives, Complexity, and Decision Time." *Econ. J.* 103 (421): 1397–417.

———. 1994. "On a Lottery Pricing Anomaly: Time Tells the Tale." *J. Risk and Uncertainty* 8 (7): 311–24.

Woodford, M. 2014. "Stochastic Choice: An Optimizing Neuroeconomic Model." *A.E.R.* 104 (5): 495–500.