

Benkert, Jean-Michel; Liu, Shuo; Netzer, Nick

Working Paper

Time is knowledge: What response times reveal

Working Paper, No. 449

Provided in Cooperation with:

Department of Economics, University of Zurich

Suggested Citation: Benkert, Jean-Michel; Liu, Shuo; Netzer, Nick (2026) : Time is knowledge: What response times reveal, Working Paper, No. 449, University of Zurich, Department of Economics, Zurich,
<https://doi.org/10.5167/uzh-261987>

This Version is available at:

<https://hdl.handle.net/10419/339359>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



**University of
Zurich**^{UZH}

University of Zurich
Department of Economics

Working Paper Series

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

Working Paper No. 449

Time is Knowledge: What Response Times Reveal

Jean-Michel Benkert, Shuo Liu and Nick Netzer

First version: August 2024
This version: February 2026

Time is Knowledge: What Response Times Reveal

Jean-Michel Benkert, Shuo Liu and Nick Netzer*

This version: February 2026

First version: August 2024

Abstract

Response times contain information about economically relevant but unobserved variables like willingness to pay, preference intensity, quality, or happiness. We provide a general characterization of the properties of latent variables that can be detected using response time data. Our theoretical framework unifies and generalizes results in the literature and gives rise to many new applications. We illustrate the rich insights that the method can deliver through several empirical applications: revealed preference analysis, identifying an optimal nudge, testing decreasing marginal happiness of income, and predicting treatment heterogeneity.

Keywords: chronometric effect, binary response model, non-parametric identification, revealed preference, nudging, decreasing marginal happiness, treatment heterogeneity.

JEL Classification: C14, D60, D91, I31

*Benkert: Department of Economics, University of Bern, jean-michel.benkert@unibe.ch. Liu: Guanghua School of Management, Peking University, and Division of Social Science, New York University Abu Dhabi, shuo.liu@gsm.pku.edu.cn. Netzer: Department of Economics, University of Zurich, nick.netzer@econ.uzh.ch. We are grateful for comments by Peter Andre, Ernst Fehr, Lukas Hensel, Blaise Melly, Lasse Mononen, Costanza Naguib, Andrew Oswald, and Martin Vaeth. We also thank seminar participants at Central University of Finance and Economics, CUHK-Shenzhen, Hong Kong Baptist University, HU Berlin, NYU Abu Dhabi, Peking University, Shanghai Jiao Tong University, Shanghai University of Finance and Economics, Xiamen University, the Universities of Bamberg, Basel, Bern, Edinburgh, Hamburg, and Hong Kong, BSE Summer Forum 2025, CESifo Area Conference on Behavioral Economics 2025, and Verein für Socialpolitik Theoretischer Ausschuss 2025. Shuo Liu acknowledges financial support from the National Natural Science Foundation of China (grants 72322006 and 72192844).

1 Introduction

Traditionally, economists have ignored choice process data like response times. Only recently has the literature realized that response times contain valuable information about unobserved variables like willingness to pay (Krajbich et al., 2012; Cotet, Zhao and Krajbich, 2025), preference intensity (Chabris et al., 2009; Alós-Ferrer, Fehr and Netzer, 2021; Alós-Ferrer and Garagnani, 2022), quality assessments (Card et al., 2024), or subjective happiness (Liu and Netzer, 2023).

In this paper, we take a systematic approach to understanding what information response times contain. We study the identification of distributional properties of latent variables in canonical binary response models, a framework extensively used in economics and applicable to all the settings described above. Our approach builds on the observation that decisions are faster when the absolute value of the latent variable is larger. The observable response times are therefore informative about the unobservable latent variable. As our main theoretical result, we provide a full characterization of the distributional properties of latent variables that can be detected using response time data. Our characterization highlights how different assumptions about the relationship between the latent variable and response times determine what properties can be identified.

To illustrate the usefulness of the method, we study four empirical applications where a diverse range of distributional properties matters. In revealed preference analysis, detecting the sign and ranking of means is crucial and enables out-of-sample predictions. For optimal nudging, the true preference distribution must be identified from choices that are distorted by frames. In the context of happiness surveys, the curvature of the income-happiness relation is of particular interest. Finally, predicting heterogeneous treatment effects requires detecting systematic group differences in latent distributions. In all cases, we show that conventional methods of analysis often suffer from identification problems, but that the use of response time data with our method offers a viable solution.

The central assumption underlying our analysis is the so-called *chronometric function* that associates each choice with a response time. This function is monotone, in the sense that a larger absolute value of the latent variable generates a faster decision, possibly after controlling for individual heterogeneity. From a theoretical perspective, such a relationship emerges naturally in evidence-accumulation models (see Chabris et al., 2009; Fudenberg, Strack and Strzalecki, 2018; Card et al., 2024), where a stronger stimulus generates faster decisions. The empirical evidence for a monotone chronometric function in the laboratory is vast. Among many others, Kellogg (1931), Moyer and Bayer (1976), and Palmer, Huk and Shadlen (2005) document the effect for choice situations with an objective stimulus, and

Moffatt (2005), Chabris et al. (2009), Konovalov and Krajbich (2019), and Alós-Ferrer and Garagnani (2022) for subjective value-based environments. Field evidence is also emerging. Card et al. (2024) document that editorial decisions take longer when the submitted paper’s quality implies a closer decision. Using eBay data on bargaining behavior, Cotet, Zhao and Krajbich (2025) show that sellers’ response times to an offer systematically depend on its perceived value. In the context of an online survey, Liu and Netzer (2023) demonstrate that faster responses are associated with a stronger sense of approval for the selected answer.

To understand our main insight, consider a decision-making environment where one or multiple agents choose between two options. Choice is determined by the realization x of an unobservable latent variable, with $x \leq 0$ generating choice of option a and $x > 0$ generating choice of option b . An analyst observing the choices wants to learn about the underlying cumulative distribution function G of the latent variable. Unfortunately, the only property of G that is identified without additional assumptions or data is its value at zero, $G(0)$, which is given by the observed probability or frequency of choosing a . This information is extremely limited and does, for example, not imply anything about the mean of G without additional distributional assumptions. Now suppose that the speed of the decision is given by $c(|x|)$ for a strictly decreasing chronometric function c , assumed here to be identical for both choice options and all subjects just for ease of exposition. Since a choice of a arises at time t or earlier if x is sufficiently far below 0, where “sufficiently far” is determined by the chronometric function, the observed probability or frequency of choosing a at time t or earlier pins down the value of $G(-c^{-1}(t))$, and analogous for choices of b . Observing the joint distribution of responses and response times therefore allows the analyst to identify a composition of the distribution G and the (inverse of the) chronometric function c .

Consequently, if the analyst had perfect knowledge of the chronometric function linking values to response times, she could recover the entire latent distribution. Such detailed knowledge is not necessary for inferring only specific distributional properties. Our main result characterizes which properties (or their violations) can be detected under which assumptions on the chronometric function. For example, detecting properties that are preserved under monotone transformations requires only knowledge of monotonicity of the chronometric function. If we further restrict attention to chronometric functions that are identical for both choice options, as in the above illustration, we can detect properties that are preserved under symmetric monotone transformations, and analogously for other classes of transformations.

Our result also provides a simple recipe how to detect or reject any property of interest. It involves constructing a candidate distribution based on the empirical data using a representative chronometric function that the analyst deems possible. If the property of interest holds for this candidate distribution, it must hold for all chronometric functions that can

be obtained from the representative one using any transformation under which the property is preserved. We discuss an extension that combines this approach with distributional assumptions. Additionally, we show how to incorporate individual heterogeneity and noise into the analysis, and we provide necessary and sufficient conditions for the rationalizability of response time data.

The existing econometrics literature has studied identification of binary response models through assumptions on the distribution of the latent variable and exogenous variation of observables (e.g., Manski, 1988; Matzkin, 1992). Some of the assumptions required to achieve identification have been criticized (e.g., Haile, Hortaçsu and Kosenok, 2008; Bond and Lang, 2019). Our method based on response times is complementary to that literature and allows us to avoid some of the controversial assumptions.

Our approach lends itself to a wide range of applications. We focus on four applications in the paper. For each of them, we use our general theoretical result to derive conditions that allow us to detect or reject distributional properties which are of interest in that specific application. We then apply these conditions to existing data and evaluate their performance.

The first application is a revealed preference analysis. In the context of stochastic choice, Alós-Ferrer, Fehr and Netzer (2021) have shown that response time data can be used to obtain revealed preferences without making distributional assumptions about the random utility component and to improve out-of-sample predictions. We first show that several of their results, such as a sufficient condition for the mean of a distribution to be positive, follow as corollaries from our characterization here. We then generalize these results, dispensing with unnecessary assumptions and allowing for heterogeneity of the chronometric effect between different decision problems. We validate these new results by showing that the out-of-sample predictions which Alós-Ferrer, Fehr and Netzer (2021) made in the data of Clithero (2018) are more accurate whenever our approach classifies them as more robust to heterogeneity.

Our second application concerns the problem of selecting an optimal nudge (Thaler and Sunstein, 2008; Benkert and Netzer, 2018). We assume that a choice architect must select between two frames, for example defaults, each distorting choice towards one of the two options. The distorted choices are not directly informative about the shares of agents who truly prefer each option. However, under the assumption that the truly indifferent agent is distorted symmetrically and chooses equally fast under the two frames, response time data allow us to detect the true preference shares and to select the optimal nudge. We apply this result to data from Serra-Garcia and Szech (2023), where the existence of a neutral frame allows us to validate the method. Our prediction is strikingly accurate and outperforms alternative methods proposed in the literature, both in the original data and in synthetic subsamples that we create to investigate the variability of the predictions.

In our third application, we test the hypothesis of decreasing marginal happiness of income, a principle central to redistributive policies. Oswald (2008) and Kaiser and Oswald (2022) question the empirical foundation of this principle by arguing that an observed concave relationship between income and self-reported happiness may result from a concave reporting function rather than from decreasing marginal happiness. Conventional approaches used in the happiness literature are insufficient to establish the principle (Bond and Lang, 2019). We show that the principle becomes testable with the response time-based method, using conditions for detecting the ranking of the means of distributions. One of these conditions is novel, another one was already derived by Liu and Netzer (2023) and once more follows as a corollary from our general result here. Our analysis of survey data by Liu and Netzer (2023) reveals that the hypothesis of decreasing marginal happiness cannot be rejected (while the hypothesis of increasing marginal happiness is clearly rejected).

Our last application concerns heterogeneous treatment effects (e.g., Manski, 2004). Heterogeneity is relevant for policymakers who need to decide which subpopulation to target with a treatment, for firms who want to set different prices in different markets, and for experimental economists who are interested in bounds on treatment effects. The existing literature typically requires post-treatment data to predict which subgroup of a population will respond more strongly to a treatment (e.g., Athey and Imbens, 2016). We show that heterogeneity can be predicted even pre-treatment, because subgroups with a larger near-indifferent mass of agents will respond more strongly to a treatment and this property can be detected using response time data. Using the data of Krefeld-Schwalb, Sugerman and Johnson (2024), we validate our approach by showing that we correctly predict the overwhelming majority of subgroup comparisons.

Our four applications are illustrations of the scope of the method and of the value of using response time data in binary response models. Going beyond these applications, the approach could also be used to detect polarization of political attitudes from ordinal survey questions (Lelkes, 2016; Vaeth, 2023), infer properties of demand functions that are important for the pricing of firms from purchase decisions at a single price (Johnson and Myatt, 2006), and uncover correlations that are not directly observable when subjects' responses are uninformative because they conflict with social norms (Coffman, Coffman and Ericson, 2017; Konovalov and Krajbich, 2019). Our Appendix D contains a collection of distributional properties that play a role in these contexts and, for each of them, derives both the identifying response time conditions and the required assumptions on the chronometric function.

The paper is organized as follows. Section 2 introduces the framework and derives our main theoretical result, followed by three extensions. Section 3 presents the four applications. Section 4 concludes. Supporting materials are provided in the appendices.

2 Theory

In this section, we develop our general theoretical framework and present our main result, which shows how and under which conditions distributional properties can be detected using response time data.

2.1 Binary Response Model

We first introduce the binary response model (e.g., Manski, 1988). There is a random variable $\tilde{x}$ with values $x \in \mathbb{R}$ that induce binary responses by comparison with a decision threshold. We normalize the threshold to zero without loss of generality. Thus, the response is $i = 0$ if $\tilde{x}$ takes a value $x \leq 0$ and $i = 1$ if $\tilde{x}$ takes a value $x > 0$. We describe the distribution of the latent variable $\tilde{x}$ by a cumulative distribution function (cdf) G , which we assume to be continuous. It follows that the probabilities of the two responses are $p^0 = G(0)$ and $p^1 = 1 - G(0)$.

The model has different applications and interpretations. For example, the latent variable could be a random utility difference $\tilde{x} = [u(1) + \tilde{\epsilon}(1)] - [u(0) + \tilde{\epsilon}(0)]$ between two options, inducing stochastic choices of a single agent (as in Alós-Ferrer, Fehr and Netzer, 2021). In another application, $\tilde{x}$ could describe the difference between the willingness to pay $\tilde{v}$ for a product among consumers and the product price p , $\tilde{x} = \tilde{v} - p$, inducing the demand for the product at that price (in the spirit of Cotet, Zhao and Krajbich, 2025). In yet another application, $\tilde{x}$ could capture the distribution of happiness in a population of agents, inducing frequencies of responses to a binary survey question about happiness (as in Liu and Netzer, 2023). The same logic applies to other survey questions, where the responses could be driven by a distribution of political or moral attitudes or preference parameters in the population. The latent variable could also describe the random quality of papers that are submitted to a journal, inducing the editor’s decision to accept or reject (as in Card et al., 2024).

We now follow Alós-Ferrer, Fehr and Netzer (2021) and Liu and Netzer (2023) and assume that the realized value x of $\tilde{x}$ not only determines the response but also the response time, with larger absolute values implying faster responses, in line with the chronometric effect.¹ Formally, we denote by $c : \mathbb{R} \rightarrow [\underline{t}, \bar{t}]$ the chronometric function, where $0 \leq \underline{t} < \bar{t} < \infty$. The function c maps each realized value x into a response time $c(x)$. It is assumed to be continuous, strictly increasing on $\mathbb{R}_-$ and strictly decreasing on $\mathbb{R}_+$ whenever $c(x) > \underline{t}$, and

¹There exists a literature in psychology that introduces response times into response models in a similar way (e.g., Takane and Sergeant, 1983; Ferrando and Lorenzo-Seva, 2007a,b; Ranger, 2013). However, that literature is interested in improving parameter estimation using response time data, rather than detecting distributional properties, and therefore imposes parametric functional forms and makes the conventional distributional assumptions.

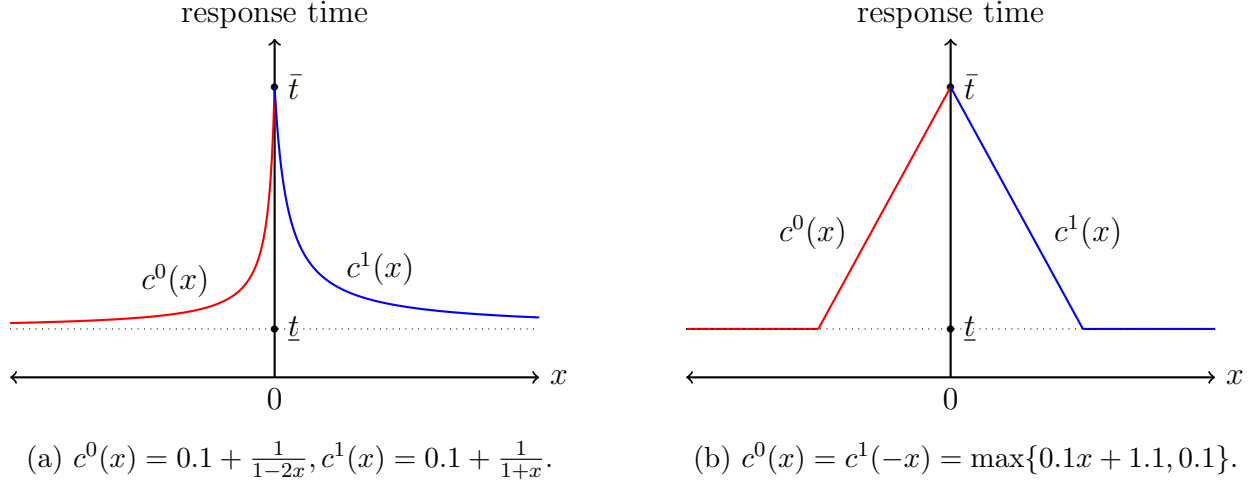


Figure 1: Examples of chronometric functions.

Notes: The left panel in the figure depicts an asymmetric chronometric function that asymptotically approaches the fastest response time $\underline{t}$, while the right panel shows a symmetric one that attains $\underline{t}$ at finite absolute values of the latent variable. In both panels, the red and blue curves correspond to the restrictions of c to $\mathbb{R}_-$ and $\mathbb{R}_+$, respectively.

to satisfy $c(0) = \bar{t}$ and $\lim_{x \rightarrow -\infty} c(x) = \lim_{x \rightarrow +\infty} c(x) = \underline{t}$. Figure 1 illustrates two examples of chronometric functions that adhere to all these conditions.

The restriction of c to $x \in \mathbb{R}_-$ is denoted c^0 . This function c^0 has a well-defined inverse $(c^0)^{-1} : (\underline{t}, \bar{t}] \rightarrow \mathbb{R}_-$ that is continuous and strictly increasing. We extend it to $\underline{t}$ by setting $(c^0)^{-1}(\underline{t}) = -\infty$ if $c(x) > \underline{t}$ for all $x \in \mathbb{R}_-$ and $(c^0)^{-1}(\underline{t}) = \max\{x \in \mathbb{R}_- \mid c(x) = \underline{t}\}$ otherwise. Analogously, the restriction of c to $x \in \mathbb{R}_+$ is denoted c^1 , with a continuous and strictly decreasing inverse $(c^1)^{-1} : (\underline{t}, \bar{t}] \rightarrow \mathbb{R}_+$, which we extend to $\underline{t}$ by setting $(c^1)^{-1}(\underline{t}) = +\infty$ or $(c^1)^{-1}(\underline{t}) = \min\{x \in \mathbb{R}_+ \mid c(x) = \underline{t}\}$, as appropriate.

In addition to response probabilities, the model (G, c) induces distributions of response times. We denote by F^i the cdf of response times conditional on a response of $i = 0, 1$. Since a response $i = 0$ at time t or earlier arises if $x \leq (c^0)^{-1}(t)$, we obtain that

$$p^0 F^0(t) = G((c^0)^{-1}(t)) \quad (1)$$

for all $t \in [\underline{t}, \bar{t}]$, where we use the convention $G(-\infty) = 0$. Analogously, a response $i = 1$ at time t or earlier arises if $x \geq (c^1)^{-1}(t)$, so that

$$p^1 F^1(t) = 1 - G((c^1)^{-1}(t)) \quad (2)$$

for all $t \in [\underline{t}, \bar{t}]$, where we use $G(+\infty) = 1$. The induced response-time cdfs F^i are continuous

on $[\underline{t}, \bar{t}]$ and satisfy $F^i(\bar{t}) = 1$. In summary, the binary response model (G, c) induces the data $(p, F) = (p^0, p^1, F^0, F^1)$ according to (1) and (2).

2.2 Detecting Properties

We now ask what an analyst can learn from observed data about the underlying binary response model and, in particular, about the distribution G of the latent variable. Taking observed data as given, different binary response models could have generated those data, so inference about the model is not straightforward. This is true especially if the analyst is not willing to make potentially strong assumptions about the form of the chronometric function or the latent distribution. We ask whether there are some properties that all models satisfy which are consistent with the data.

Consider a profile of data $(p_j, F_j)_j = (p_j^0, p_j^1, F_j^0, F_j^1)_j$ indexed by $j \in J$. The set J could be a singleton, e.g., when studying the choices of a single agent between two options or the responses of a single group of agents to one binary survey question. In this case, we typically omit the index j . If we observe the choices of a single agent for multiple pairs of options, then the set J would describe the different binary choice problems. Alternatively, the index could capture different agents or combinations of agents and choice problems. In a survey application, the index could describe different questions or different demographic groups. Since J is not necessarily finite, it could also model income levels $j \in J \subseteq \mathbb{R}_+$ of survey participants. In the other applications discussed earlier, the index could describe different journal editors or different prices at which market demand is observed. We assume that each F_j^i is continuous and satisfies $F_j^i(\underline{t}) = 0$ and $F_j^i(\bar{t}) = 1$.²

Denote by $\mathcal{G}$ the set of all possible profiles $(G_j)_j$ of cdfs, where each individual G_j satisfies our assumptions from the previous subsection. Similarly, denote by $\mathcal{C}$ the set of all possible profiles $(c_j)_j$ of chronometric functions that individually satisfy our previous assumptions. For our main result, we leave $\mathcal{G}$ unrestricted but allow for a possibly restricted set $\mathcal{C}^* \subseteq \mathcal{C}$ of admissible chronometric functions. For example, each profile $(c_j)_j \in \mathcal{C}^*$ may have to satisfy that all functions are symmetric across responses (i.e., $c_j^0(-x) = c_j^1(x)$ for all $x \in \mathbb{R}_+$). This restriction embodies the assumption that the chronometric effect is identical for the two choice options. Another possible restriction would be that the functions are identical across indices (i.e., $c_j = c$ for all $j \in J$), reflecting the assumption that the chronometric effect is the same for all groups j . Other restrictions could be functional forms such as piece-wise linearity, or combinations of multiple of these assumptions.

²Even though our approach is very general, it does not allow describing correlations between choices or response times across different indices j . We leave an extension that allows for such correlations across decision problems for future work.

Definition 1. Given data $(p_j, F_j)_j$ and a set $\mathcal{C}^*$ of admissible chronometric functions, a property **P** of the distributions $(G_j)_j$ is *detected* if all $((G_j)_j, (c_j)_j) \in \mathcal{G} \times \mathcal{C}^*$ that induce $(p_j, F_j)_j$ —in the sense of (1) and (2) for all $j \in J$ —have in common that $(G_j)_j$ satisfies **P**.

This idea of detection formalizes that property **P** is inferred strictly from the data instead of being derived based on assumptions by the analyst. It corresponds to a standard notion of non-parametric identification in econometrics (e.g., Manski, 1988; Bond and Lang, 2019). The revealed preference approach in choice theory (Samuelson, 1938; Arrow, 1958) embodies the same logic (see, e.g., Benkert and Netzer, 2018; Alós-Ferrer, Fehr and Netzer, 2021).³

The ability to detect a property **P** will depend on the extent to which **P** is invariant to transformations. For example, the property that a single cdf G is strictly increasing (full support) is invariant to all strictly increasing transformations ψ , because $G(\psi(x))$ is still strictly increasing in x when the function ψ is strictly increasing. As another example, the property that G_1 first-order stochastically dominates G_2 is invariant to profiles (ψ_1, ψ_2) of strictly increasing transformations that are identical for the two distributions (i.e., $\psi_1 = \psi_2$), but not to distribution-specific transformations. Generally, let Ψ be a set of profiles $(\psi_j)_j$ of functions $\psi_j : \mathbb{R} \rightarrow \mathbb{R}$ that are bijective and strictly increasing, hence continuous. The set Ψ can embody various constraints, such as the restriction that each profile $(\psi_j)_j \in \Psi$ is composed of identical functions (i.e., $\psi_j = \psi$ for all $j \in J$). For any $(G_j)_j \in \mathcal{G}$ and any $(\psi_j)_j \in \Psi$, the composition $(G_j \circ \psi_j)_j$ is another profile of cdfs in $\mathcal{G}$.

Definition 2. A property **P** of $(G_j)_j$ is *invariant to transformations* Ψ if $(G_j \circ \psi_j)_j$ also has property **P**, for all $(\psi_j)_j \in \Psi$.

Classifying properties by their invariance to transformations is common in the theory of measurement (Luce et al., 1990) and parallels an approach used in social choice theory to describe the measurability and interpersonal comparability of utilities that different welfare functions require (d’Aspremont and Gevers, 2002).

2.3 Generating Chronometric Functions

We consider sets $\mathcal{C}^*$ of chronometric functions that are generated by a representative profile $(c_j^*)_j \in \mathcal{C}$ and a set of transformations Ψ as introduced above.

Definition 3. The pair $((c_j^*)_j, \Psi)$ *generates* $\mathcal{C}^*$ if for each $(c_j)_j \in \mathcal{C}^*$ there exists $(\psi_j)_j \in \Psi$ such that $(c_j)_j = (c_j^* \circ \psi_j)_j$.

³We assume here implicitly that there exists at least one $((G_j)_j, (c_j)_j) \in \mathcal{G} \times \mathcal{C}^*$ that induces the data. As we will see later, this is always the case when $\mathcal{G}$ is unrestricted. In Subsection 2.5.2 we will study the case with restricted $\mathcal{G}^* \subseteq \mathcal{G}$ and provide conditions for rationalizability of the data.

Note that only transformations ψ_j that satisfy $\psi_j(0) = 0$ yield well-defined chronometric functions, while Ψ could contain functions without that property. More generally, the definition does not require that all elements of Ψ must be used for the construction of $\mathcal{C}^*$.

To illustrate the concept, we discuss examples of sets $\mathcal{C}^*$ and how they can be generated. Consider the set of all profiles of chronometric functions which approach $\underline{t}$ asymptotically in the limit but never reach $\underline{t}$. This set is generated by a simple representative member $(c_j^*)_j$, for example the symmetric hyperbolic form

$$c_j^*(x) = \underline{t} + \frac{1}{|x| + 1/(\bar{t} - \underline{t})}, \quad (3)$$

identical for each $j \in J$, together with the unrestricted set of all profiles of transformations. To see why, just note that any desired $(c_j)_j$ can be obtained from $(c_j^*)_j$ by using the transformations $(\psi_j)_j$ given by

$$\psi_j(x) = \begin{cases} (c_j^{*,1})^{-1}(c_j(x)) & \text{if } x > 0, \\ (c_j^{*,0})^{-1}(c_j(x)) & \text{if } x \leq 0, \end{cases}$$

for each $j \in J$. Similarly, the set of all profiles of chronometric functions which have $c_j(x) = \underline{t}$ for large but finite absolute values of x can be generated by a simple representative member, for example the symmetric linear form

$$c_j^*(x) = \begin{cases} \underline{t} & \text{if } (\bar{t} - \underline{t}) < x, \\ \bar{t} - x & \text{if } 0 < x \leq (\bar{t} - \underline{t}), \\ \bar{t} + x & \text{if } -(\bar{t} - \underline{t}) \leq x \leq 0, \\ \underline{t} & \text{if } x < -(\bar{t} - \underline{t}), \end{cases} \quad (4)$$

identical for each $j \in J$, together with the unrestricted set of all profiles of transformations.⁴

If we combine either (3) or (4) with the smaller set of transformations that are symmetric around zero ($\psi_j(-x) = -\psi_j(x)$ for all $x \in \mathbb{R}_+$), we can generate the respective sets of profiles of chronometric functions that are symmetric across responses. If we combine (3) or (4) only with transformations that are identical across indices, we generate only profiles of chronometric functions that are identical across indices. Different of these cases will be relevant in our applications in Section 3.

⁴An example of a set $\mathcal{C}^*$ which cannot be generated by any $((c_j^*)_j, \Psi)$ is one which contains some profiles in which c_j reaches $c_j(x) = \underline{t}$ and other profiles in which $c_j(x) > \underline{t}$ throughout, for some $j \in J$. However, this can be addressed by our generalization in Subsection 2.5.1 which allows for the generation of sets by more than one representative profile $(c_j^*)_j$.

2.4 Main Result

Given observed data $(p_j, F_j)_j$ and a representative profile $(c_j^*)_j$ of chronometric functions, we can derive the empirical distribution functions

$$H_j(x) = \begin{cases} 1 - p_j^1 F_j^1(c_j^*(x)) & \text{if } x > 0, \\ p_j^0 F_j^0(c_j^*(x)) & \text{if } x \leq 0, \end{cases} \quad (5)$$

for all $j \in J$. Each H_j is a well-defined and continuous cdf. It would be the true cdf of the binary response model inducing (p_j, F_j) if c_j^* was the true chronometric function.⁵

Of course, c_j^* may be different from the true chronometric function. However, the proof of our main result shows that whenever an admissible chronometric function c_j can be written as $c_j = c_j^* \circ \psi_j$ for some ψ_j , the corresponding distribution G_j can be written as $G_j = H_j \circ \psi_j$. Therefore, if $(H_j)_j$ satisfies a property that is invariant to Ψ , then any $(G_j)_j$ that is compatible with the data also satisfies that property, under the assumption that $((c_j^*)_j, \Psi)$ generates $\mathcal{C}^*$. This gives rise to the following result.

Theorem 1. *Suppose $\mathcal{C}^*$ is generated by $((c_j^*)_j, \Psi)$. If $(H_j)_j$ defined in (5) satisfies a property $\mathbf{P}$ that is invariant to transformations Ψ , then $\mathbf{P}$ is detected.*

Proof. Let $((G_j)_j, (c_j)_j) \in \mathcal{G} \times \mathcal{C}^*$ be any model that induces $(p_j, F_j)_j$. Using the definition of $(H_j)_j$ and the conditions (1) and (2) for inducing the data, we obtain that

$$H_j(x) = \begin{cases} G_j((c_j^1)^{-1}(c_j^{*1}(x))) & \text{if } x > 0, \\ G_j((c_j^0)^{-1}(c_j^{*0}(x))) & \text{if } x \leq 0, \end{cases}$$

for each $j \in J$. Since $((c_j^*)_j, \Psi)$ generates $\mathcal{C}^*$, there exists $(\psi_j)_j \in \Psi$ such that $c_j = c_j^* \circ \psi_j$, for each $j \in J$. We therefore have $c_j^0(x) = c_j^{*0}(\psi_j(x))$ for any $x \leq 0$ and $c_j^1(x) = c_j^{*1}(\psi_j(x))$ for any $x > 0$. It follows that $(c_j^0)^{-1}(t) = \psi_j^{-1}((c_j^{*0})^{-1}(t))$ and $(c_j^1)^{-1}(t) = \psi_j^{-1}((c_j^{*1})^{-1}(t))$ for all $t \in [\underline{t}, \bar{t}]$. It then follows that

$$H_j(x) = \begin{cases} G_j(\psi_j^{-1}(x)) & \text{if } x > 0, \\ G_j(\psi_j^{-1}(x)) & \text{if } x \leq 0. \end{cases}$$

Note that this is also true when $\bar{x}_j^* := (c_j^{*0})^{-1}(\underline{t})$ is finite and we consider any $x \leq \bar{x}_j^*$. In that case, $\bar{x}_j := (c_j^0)^{-1}(\underline{t}) = \psi_j^{-1}(\bar{x}_j^*)$ is also finite, and we have $G_j(x) = 0$ for all $x \leq \bar{x}_j$ because (G_j, c_j) induces data with $F_j^0(\underline{t}) = 0$. For all $x \leq \bar{x}_j^*$, we therefore have $G_j((c_j^0)^{-1}(c_j^{*0}(x))) =$

⁵See Alós-Ferrer, Fehr and Netzer (2021, Online Appendix p. 5) for an analogous construction.

$G_j(\bar{x}_j) = 0 = G_j(\psi_j^{-1}(\bar{x}^*)) = G_j(\psi_j^{-1}(x))$. The analogous argument applies to $x \geq (c_j^*)^{-1}(\underline{t})$. To summarize, we have $H_j = G_j \circ \psi_j^{-1}$, for each $j \in J$. If $(H_j)_j$ satisfies a property $\mathbf{P}$ that is invariant to transformations Ψ , then $(H_j \circ \psi_j)_j = (G_j \circ \psi_j^{-1} \circ \psi_j)_j = (G_j)_j$ also satisfies $\mathbf{P}$, hence $\mathbf{P}$ is detected. $\square$

As we will show in Section 3, Theorem 1 is easy to apply in a broad range of applications. If we are interested in some distributional property, we only need to check specific empirical distributions $(H_j)_j$ that are based on a representative profile $(c_j^*)_j$ of chronometric functions. These distributions can be constructed from the observed data. If they satisfy a property that is invariant to transformations Ψ , then this property is detected for the entire class of chronometric functions that $((c_j^*)_j, \Psi)$ generates. Often, the resulting detection conditions can be expressed directly and transparently in terms of the observable data, without recourse to specific functions $(c_j^*)_j$.

The theorem can also be used to detect whether a property is violated. Denote by $\neg\mathbf{P}$ the property that property $\mathbf{P}$ is not true. The theorem immediately implies that if the empirical distributions $(H_j)_j$ satisfy $\neg\mathbf{P}$ and this property is invariant to transformations Ψ , then $\neg\mathbf{P}$ is detected. In words, the violation of a property is detected when $(H_j)_j$ violates the property and the violation is invariant to transformations Ψ .⁶

2.5 Extensions

2.5.1 Multiple Representative Chronometric Functions

Our first extension covers the case where $\mathcal{C}^*$ is not generated from one representative profile of chronometric functions $(c_j^*)_j$ but from multiple profiles $(c_j^{k*})_j$ where $k \in K$. We say that $((c_j^{k*})_j^k, \Psi)$ generates $\mathcal{C}^*$ if for each $(c_j)_j \in \mathcal{C}^*$ there exist $(\psi_j)_j \in \Psi$ and $k \in K$ such that $(c_j)_j = (c_j^{k*} \circ \psi_j)_j$. For example, if we want to generate a function c_j that never reaches $\underline{t}$, we need to start from a representative function c_j^{1*} with that same property, like (3). If we simultaneously want to generate another function that has $c_j(x) = \underline{t}$ for large absolute values of x , we need a different representative function c_j^{2*} with that property, like (4). Another example, which we will encounter in our applications in Section 3, is the use of multiple representative profiles to capture the analyst's uncertainty about the asymmetry of the chronometric function between the options or its heterogeneity across

⁶Note that $\mathbf{P}$ and $\neg\mathbf{P}$ can be invariant to the same or to different sets of transformations. We also remark that detecting $\neg\mathbf{P}$ is stronger than not detecting $\mathbf{P}$. A property is not detected if *at least one* model that is compatible with the data violates the property, while detecting $\neg\mathbf{P}$ requires that *all* models that are compatible with the data violate the property. See Liu and Netzer (2023, p. 3303) for a discussion of this difference in the context of first-order stochastic dominance of happiness distributions.

groups. Working with multiple representative functions that capture different degrees of asymmetry or heterogeneity allows to robustly detect properties or establish effect bounds.

Instead of a single empirical profile $(H_j)_j$, we obtain one profile $(H_j^k)_j$ for each $k \in K$, defined as in (5) using the respective functions $(c_j^{k*})_j$. Theorem 1 then becomes that property $\mathbf{P}$ is detected if $(H_j^k)_j$ satisfies $\mathbf{P}$ for each $k \in K$ (under the otherwise identical assumptions). In words, we simply need to repeat the previous procedure for each $k \in K$ separately, and we achieve detection if we achieve it for each $k \in K$.

2.5.2 Distributional Assumptions

Our second extension covers the case where the set of admissible distributions is restricted to some $\mathcal{G}^* \subseteq \mathcal{G}$. Any such restriction reflects prior knowledge of the analyst about properties of the distributions (e.g., Manski, 1988). For example, in an application where each G_j describes the willingness to pay of consumers minus the corresponding product price, the different distributions in the profile $(G_j)_j$ should all be horizontal shifts (by the difference in product prices) of one identical distribution.

Knowledge about the distributions can make detection easier. We will exploit this in one of our applications in Section 3, where we want to detect whether the mean of a distribution G_1 is larger than the mean of a distribution G_2 . This property is invariant to identical positive affine transformations but not to identical monotone transformations in general. Hence, detecting it requires relatively strong assumptions on the set of admissible chronometric functions. Consider instead the property that the median of G_1 is larger than the median of G_2 . This property is invariant to all identical monotone transformations and can therefore be detected under milder conditions. We can compensate these milder conditions by assuming that all distributions in $\mathcal{G}^*$ are symmetric around their means, an assumption frequently made in binary response models. Since mean and median coincide in that case, a detected ranking of the medians implies a detected ranking of the means.

A question that arises when working with restrictions on both $\mathcal{G}^*$ and $\mathcal{C}^*$ is whether the given data are *rationalizable*, i.e., whether there exist distributions and chronometric functions in the restricted sets that induce the data.⁷ In Appendix A we study this question and provide necessary and sufficient conditions for rationalizability.

⁷Echenique and Saito (2017) study rationalizability for the case of deterministic responses and response times. Alós-Ferrer, Fehr and Netzer (2021) discuss the difficulty of the problem when responses and response times are stochastic. For a single binary choice problem, they show that all stochastic choice functions with response times are rationalizable when the distributions are unrestricted. For multiple binary choice problems, they discuss necessary conditions for rationalizability as well as the possibility of rationalization by generalized random utility models.

2.5.3 Heterogeneity and Noise

Before applying our method to data, it is natural to ask about individual heterogeneity in response speed and about noise or measurement error (see, e.g., Johnson, 2004). Alós-Ferrer, Fehr and Netzer (2021) and Liu and Netzer (2023) show that several of their results have appropriate extensions that facilitate empirical testing. We follow the same approach and model the response time of an individual with realized latent value x by $t = c_j(x) \cdot \eta \cdot \epsilon$, where $c_j(x)$ is the chronometric function, $\eta > 0$ captures the individual’s general response speed, and $\epsilon > 0$ is a noise term comprising uncontrolled factors and measurement error. We assume that the terms (x, η) are drawn from a joint distribution, possibly j -specific, with marginal cdf $G_j(x)$ whose properties we are interested in. Moreover, we assume that the analyst has access to a baseline response, such as a demographic survey question, where the response time is given by $t_b = \phi \cdot \eta$ with intensity parameter $\phi > 0$. The terms (ϵ, ϕ) are also drawn from a joint distribution, possibly j -specific, but independent of the other variables.

The baseline response allows us to normalize response times in the decision problem of interest (see Fazio, 1990). Working with $\hat{t} = t/t_b = c_j(x) \cdot \epsilon/\phi$ takes care of the individual heterogeneity embodied in η . However, the analyst who uses such normalized data $(p_j, \hat{F}_j)_j$ to derive candidate distributions $(\hat{H}_j)_j$ according to (5) is still misspecified, as she ignores the variation embodied in ϵ and ϕ .

In Appendix B, we derive conditions under which the distributions $(\hat{H}_j)_j$ still inherit a property **P** from $(G_j)_j$ and can therefore be used for testing the null hypothesis that **P** is satisfied by the true distributions. In addition to invariance properties similar to our main analysis, the key requirement is invariance to multiplicative convolutions of the form

$$\hat{G}_j(x) = \int G_j(xz) d\mu_j(z).$$

This is satisfied by all the properties that we will use for hypothesis testing in Section 3, and more broadly by all properties that reduce to “pointwise arguments” in the integral.⁸

3 Applications

In this section, we study four applications and, in each of them, show how our theoretical result can be used to detect or reject economically meaningful distributional properties.

⁸There are also properties for which these convolutions are not innocuous. It is known, for instance, that single-crossing of cdfs does not aggregate in general (Quah and Strulovici, 2012). When attempting to test such a condition in noisy data, one may have to look for restricted classes of latent distributions and noise terms under which testing remains possible.

3.1 Revealed Preference Analysis

As a first application, we use response time data for revealed preference analysis. Consider a random utility model where $\tilde{x} = [u(a) + \tilde{\epsilon}(a)] - [u(b) + \tilde{\epsilon}(b)]$ and the errors $\tilde{\epsilon}(\cdot)$ have mean zero. The mean of $\tilde{x}$ equals the utility difference $u(a) - u(b)$, so that deducing the agent’s ordinal preference between a and b can be rephrased as detecting the sign of the mean.⁹

The sign of the mean is invariant to increasing linear transformations. Detecting it thus requires knowledge of the chronometric function up to such positive rescaling, which is demanding. To achieve more robust detection, we can work with a property that is sufficient for a positive (or negative) mean. Consider the asymmetry property that $G(-x) \leq 1 - G(x)$ for all $x \in \mathbb{R}_+$. This property implies that the mean of G is positive (see Alós-Ferrer, Fehr and Netzer, 2021). It is invariant to all transformations that are symmetric around zero but not necessarily linear. Detecting it thus requires knowledge of the chronometric function only up to symmetric transformations. The most natural assumption guaranteeing this is that the chronometric function is symmetric, i.e., that the chronometric effect is identical for the two options. Then, irrespective of which symmetric c^* we use, the distribution H defined in (5) exhibits the desired asymmetry if and only if

$$p^b F^b(t) \leq p^a F^a(t) \text{ for all } t \in [\underline{t}, \bar{t}]. \quad (6)$$

Inequality (6) is expressed directly in terms of the observed data and is a sufficient condition for a revealed preference $u(b) \leq u(a)$. An analogous condition holds for strict preferences.

This inequality is the same as in Theorem 1 of Alós-Ferrer, Fehr and Netzer (2021).¹⁰ As these authors discuss in detail, observing choice frequencies $p^b \leq p^a$ is not sufficient for a preference $u(b) \leq u(a)$ to be revealed without the additional assumptions on error distributions of conventional logit or probit models, posing a severe identification problem for random utility models. However, if the inequality holds for all response times, as stated in (6), then a preference is robustly revealed without distributional assumptions. They report that slightly more than 60% of all stochastic choices in the data of Clithero (2018) do robustly reveal a preference. Several other authors have by now used the same condition to answer open questions in different settings. Alós-Ferrer, Garagnani and Fehr (2024) show that a sizable fraction of choices that violate stochastic transitivity in different experiments

⁹Some distributions do not have a mean. This can be dealt with either by using the procedure described in Subsection 2.5.2 to restrict the set of distributions to those which have a mean, or by refining the desired property, for example to “the mean exists and is positive.” Analogous arguments apply whenever a property is not well-defined for all possible distributions.

¹⁰Alós-Ferrer, Fehr and Netzer (2021) allow all chronometric functions that are symmetric. One minor difference is that they assume the chronometric function to be unbounded while we assume that $\bar{t}$ is finite.

reveal non-transitive preferences and thus cannot be explained by transitive preferences together with noise. Similarly, Alós-Ferrer et al. (2024) show that common ratio and common consequence effects reflect preferences rather than noise, and Castillo and Vitaku (2025) confirm that choice reversals in line with salience theory reflect true preference reversals.

Now consider a setting with multiple binary choice problems. Suppose we observe an agent's choices between the two options a and z and between the two options b and z . Can we infer the agent's preference between a and b from these choices? Sometimes this is possible based on transitivity of preferences, for example if a is chosen over z and z is chosen over b . If, by contrast, both a and b are chosen over z , then we cannot rank a and b directly. Krajbich, Oud and Fehr (2014) noted, however, that the preference can be deduced from response times under the assumption of a monotone chronometric effect. When the choice of a over z is faster than the choice of b over z , then $u(a) - u(z)$ must be larger than $u(b) - u(z)$ and we can conclude that $u(b) \leq u(a)$ (see also Echenique and Saito, 2017). Alós-Ferrer, Fehr and Netzer (2021) adapted this argument for stochastic choice. Suppose that the random utility difference between a and z is described by a cdf G_{az} with mean $u(a) - u(z)$, and the random utility difference between b and z is described by G_{bz} with mean $u(b) - u(z)$. Deducing a revealed preference for a over b can be rephrased as detecting that the mean of G_{az} is larger than the mean of G_{bz} .

Let us now follow Alós-Ferrer, Fehr and Netzer (2021) and assume that G_{az} and G_{bz} are symmetric around their means. Mean and median coincide in this case, and hence we can try to detect whether the median of G_{az} is larger than the median of G_{bz} . The ranking of medians is invariant to all transformations that are identical for the two distributions, so that we need to know the chronometric functions only up to identical transformations. A natural assumption guaranteeing this is that the chronometric effect is identical in the two binary choice problems but otherwise unrestricted. Consider then the case $p_{az}^a > 1/2$ and $p_{bz}^b > 1/2$, where lower indices describe the choice problem and upper indices describe the chosen option (the case where $p_{az}^a < 1/2$ and $p_{bz}^b < 1/2$ is analogous). Under the assumption of symmetric distributions, this implies that both $u(a) - u(z)$ and $u(b) - u(z)$ are strictly positive, so that a preference between a and b does not follow from transitivity. Define θ_{az} and θ_{bz} as specific percentiles of the response time distributions when a or b were chosen over z , respectively:

$$F_{az}^a(\theta_{az}) = \frac{1}{2p_{az}^a} \quad \text{and} \quad F_{bz}^b(\theta_{bz}) = \frac{1}{2p_{bz}^b}.$$

It is now an easy exercise to show that the median of H_{az} is larger than the median of H_{bz} ,

where both distributions are constructed by (5) based on an identical c^* , if and only if

$$\theta_{az} \leq \theta_{bz}. \quad (7)$$

Inequality (7) is again expressed in terms of the observed data and implies the out-of-sample revealed preference $u(b) \leq u(a)$. An analogous statement holds for strict preferences.

Inequality (7) is stated in Theorem 2 of Alós-Ferrer, Fehr and Netzer (2021). As they discuss in detail, symmetry of the distributions is not sufficient to learn the out-of-sample preference between a and b from the choice probabilities p_{az}^a and p_{bz}^b . However, (7) formalizes the notion that the choice of a over z is faster than the choice of b over z and implies the preference. They then use the result to make out-of-sample predictions in the data of Clithero (2018) and show that about 80% of the predictions are correct, as compared to only about 74% for a conventional logit model. Alós-Ferrer and Garagnani (2024) have subsequently used the approach to predict subjects' choices under risk in different experiments. They show that it outperforms the predictions of various structural models of behavior.

Our paper adds to these existing results in several ways. First, it unifies them by showing that they can be derived as special cases from our Theorem 1. Second, it clarifies the intuition underlying (7) as a condition which detects the ranking of two medians. Third, it reveals that some assumptions made in the earlier literature are not necessary. For example, the assumption of a symmetric chronometric function in Alós-Ferrer, Fehr and Netzer (2021) is not required for the validity of (7). Finally, it paves the way for generalizations. We will illustrate this by showing how the above out-of-sample predictions can be generalized to accommodate differences in chronometric functions between the binary choice problems.

If we have reason to believe that the chronometric effect is heterogeneous across choice problems, we can construct H_{az} and H_{bz} based on different representative chronometric functions c_{az}^* and c_{bz}^* . Specifically, let those functions be $c_{az}^* = c^*$ and $c_{bz}^* = m \circ c^*$ for some adjustment function $m : [\underline{t}, \bar{t}] \rightarrow [\underline{t}, \bar{t}]$ that reflects our prior knowledge of the differences. Then the condition analogous to (7) reads

$$\theta_{az} \leq m^{-1}(\theta_{bz}). \quad (8)$$

For example, if we know that choices of b are faster than choices of a for any fixed utility difference to z (maybe because b and z are easier to compare than a and z , see Gonçalves, 2024), then m^{-1} makes the right hand side of (8) larger and the inequality easier to satisfy. The opposite is true when the choices of a are faster. In the plausible case where we expect differences but without knowing details, we can construct multiple functions H_{az}^k and H_{bz}^k using different adjustments m^k that reflect varying degrees of heterogeneity, as outlined in

Section 2.5.1. To obtain an out-of-sample revealed preference $u(b) \leq u(a)$, the modified conditions (8) then have to hold simultaneously for all k , ultimately reducing to the requirement that the gap between θ_{az} and θ_{bz} must be large enough.

We now apply this new result to study the robustness of the predictions by Alós-Ferrer, Fehr and Netzer (2021). Using the experimental data of Clithero (2018), they predicted the choices of 31 subjects for pairs of 17 snacks. The prediction is out-of-sample because it rests only on data from choices between each of the 17 snacks and one common reference snack, which takes the role of z from our above line of argument. Alós-Ferrer, Fehr and Netzer (2021) constructed a complete preference order for each subject, using the response time-based approach described above. Then they predicted the higher-ranked option to be chosen in each pair of snacks and computed the fraction of correct predictions. Here we consider how that fraction of correct predictions changes when restricting attention to pairs (a, b) for which the gap between θ_{az} and θ_{bz} is sufficiently large. If there is heterogeneity in chronometric effects between pairs, revealed preferences should be more robust and predictions should be more accurate when the gap is larger.

Figure 2 plots the fraction of correct predictions (on the y-axis) when restricting attention to pairs for which the absolute difference $|1/\theta_{az} - 1/\theta_{bz}|$ is larger than any given value δ (on the x-axis).¹¹ The dashed blue line refers to all predictions where transitivity is not applicable and which are therefore exclusively based on a comparison of the response time percentiles. For $\delta \rightarrow 0$, the accuracy of these predictions approaches an overall value of about 65%, clearly better than chance. As δ grows, it increases strongly and eventually reaches 100%. The solid black line refers to all predictions, including those where transitivity is applicable. The accuracy approaches 80% for $\delta \rightarrow 0$, which is the respective value reported in Alós-Ferrer, Fehr and Netzer (2021). It again grows in δ and approaches 100%. The figure also plots as a dotted orange line the accuracy for the predictions that are based on transitivity alone. These predictions do not require response times, so the difference between the percentiles here does not reflect robustness to heterogeneity in the chronometric effect. The line is essentially flat, compatible with our interpretation that the generalized revelation condition (8) captures robustness to chronometric heterogeneity.

Overall, the analysis suggests the presence of some heterogeneity in chronometric effects between pairs. Hence, using our robust condition (8) is warranted. In particular, by explicitly

¹¹More precisely, we use the difference $|1/\theta_{az} - 1/\theta_{bz}|$ when $p_{az}^a > 1/2$ and $p_{bz}^b > 1/2$. When $p_{az}^a < 1/2$ and $p_{bz}^b < 1/2$, we use $|1/(-\theta_{za}) - 1/(-\theta_{zb})|$, where the percentiles are defined as earlier but with flipped roles of the options. When $p_{az}^a > 1/2 > p_{bz}^b$, so that a revealed preference follows from transitivity under our assumption of symmetric distributions, we work with $|1/(\theta_{az}) - 1/(-\theta_{zb})|$. In the few cases where a choice probability is exactly 1/2, we set the inverse of the percentile equal to zero in these expressions. We derive all percentiles using empirical CDFs without smoothing, see Alós-Ferrer, Fehr and Netzer (2021) for details.

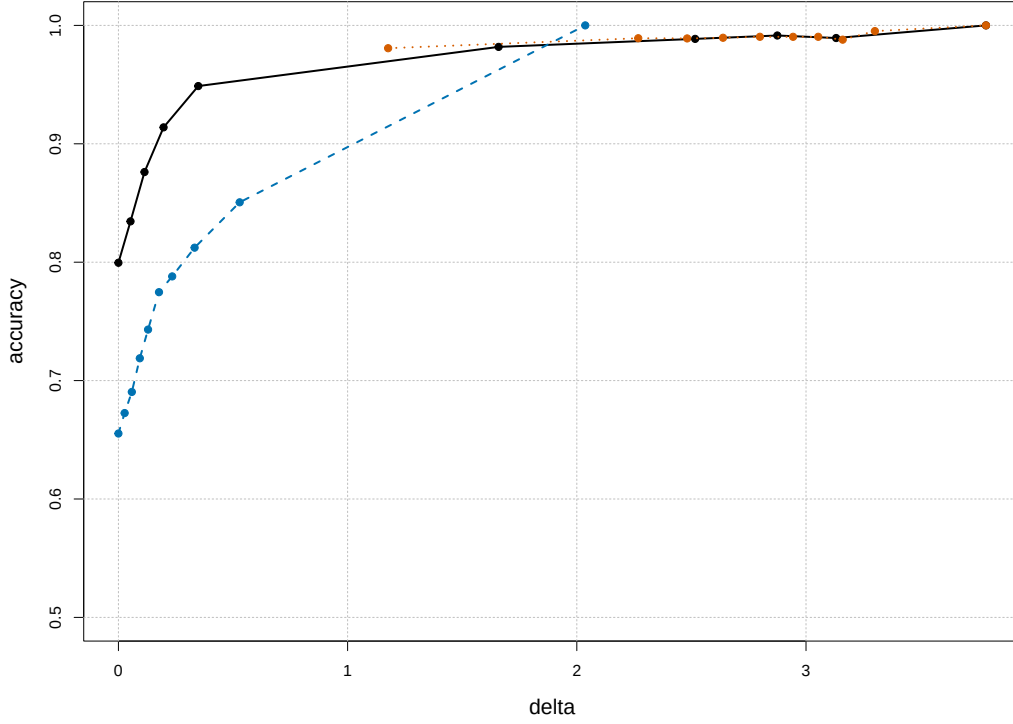


Figure 2: Prediction accuracy with chronometric heterogeneity

Notes: The curves depict the fraction of correct out-of-sample predictions on the y-axis when restricting attention to pairs with a percentile gap at least as large as the value on the x-axis. The solid black line refers to all pairs, the dashed blue line refers to only those pairs where transitivity makes no prediction, and the dotted orange line refers to pairs where transitivity alone pins down the prediction. The data are binned so that each dot represents an equal number of observations.

allowing for heterogeneity in chronometric functions, our framework yields sharper and more reliable revealed-preference predictions than previous approaches.

3.2 Identifying an Optimal Nudge

Choices are often affected by the framing of a decision problem. Examples include the role of defaults for 401(k) savings (Madrian and Shea, 2001), the behavioral effect of phrasing the consequences of a disease as either gains or losses (Tversky and Kahneman, 1981), and the relevance of the order of presentation for food choices (Bucher et al., 2016). The literature on nudging has promoted the idea of exploiting such framing effects to influence choices in a non-coercive manner (Thaler and Sunstein, 2003, 2008). An important question arising in that case is how to identify the agents' true preferences and the optimal nudge (Benkert and Netzer, 2018). Given that observed choices are necessarily distorted by framing, how can it

be justified that a frame like a savings default helps agents make decisions that are more in line with their true preferences?

Response time data may help to answer this question. To see how, consider the following simple model of choice with frames. The true and non-distorted distribution of the latent variable in a heterogeneous population of agents is given by G . This variable could be a preference parameter measuring patience, a political or moral attitude, or the willingness to pay for a product. The decision problem is necessarily framed in a way that creates a bias in favor of one of the two responses. Choices under frame f_+ are generated by a distribution G_+ that is biased in favor of $i = 1$, and choices under frame f_- are generated by a distribution G_- that is biased in favor of $i = 0$. We impose only one assumption on the relation between G and its biased versions, namely that there exists $z \geq 0$ with $G_-(-z) = G(0) = G_+(+z)$. This condition holds, for instance, when framing induces a symmetric distortion of the true indifference point without altering its ordinal position in the distribution. In that case, an agent who is truly indifferent between the two options perceives them as having a utility difference of $-z$ or $+z$, respectively. The value of z is not assumed to be known. Beyond the above condition, the shapes of G_- and G_+ can be distorted arbitrarily from G . With only this assumption, it is impossible to learn the non-distorted fraction $p^0 = G(0)$ of agents who truly prefer $i = 0$, when observing only the distorted fractions $p_-^0 = G_-(0)$ and $p_+^0 = G_+(0)$. Knowing p^0 , however, is crucial in a range of applications. For example, we need to know p^0 to identify which of the two frames induces the optimal choice for the majority of agents.

For the pair (G_-, G_+) of distributions, the property that we want to detect is hence the value p^0 such that $p^0 = G_-(-z) = G_+(+z)$ for some $z \geq 0$. Note that this value p^0 is unique for a given pair of distributions even though z may not be. However, different pairs of distributions that are compatible with the same data could exhibit different values of p^0 . The exact value of p^0 is invariant to all transformations (ψ_-, ψ_+) of (G_-, G_+) satisfying $\psi_-(-x) = \psi_+(x)$ for all $x \geq 0$. That is, the restrictions of ψ_- to $\mathbb{R}_-$ and of ψ_+ to $\mathbb{R}_+$ coincide after a change of sign. A natural assumption guaranteeing that the chronometric functions are known up to these transformations is that the chronometric effect when choosing the option favored by the frame is identical across frames.¹² Given arbitrary representative chronometric functions with that property, we can then construct distributions (H_-, H_+) by (5), and Theorem 1 implies that we detect the value p_{RT}^0 for which

$$p_{RT}^0 = p_-^0 F_-^0(t^*) = 1 - p_+^1 F_+^1(t^*) \quad (9)$$

holds for some $t^* \in [\underline{t}, \bar{t}]$. Intuitively, a truly indifferent agent is equally fast choosing $i = 0$

¹²This allows the frame to have a direct impact on response times, for example when choosing the default requires fewer actions and hence less time, as long as this does not depend on which option is the default.

under frame f_- as choosing $i = 1$ under frame f_+ . The mass of agents who truly prefer $i = 0$ therefore equals both the mass of agents who under frame f_- choose $i = 0$ earlier than the indifferent agent and the mass of agents who under frame f_+ do not choose $i = 1$ earlier than the indifferent agent. Condition (9) determines the response time of the truly indifferent agent by equating these two masses. The term $p_-^0 F_-^0(t)$ is a continuous and weakly increasing function that starts at 0 for $t = \underline{t}$ and ends at p_-^0 for $t = \bar{t}$. The term $1 - p_+^1 F_+^1(t)$ is a continuous and weakly decreasing function which starts at 1 and ends at $1 - p_+^1 = p_+^0 \leq p_-^0$. Hence, while the value t^* that solves (9) is not necessarily unique, the detected value p_{RT}^0 is unique and satisfies $p_+^0 \leq p_{RT}^0 \leq p_-^0$. Response times allow us to determine where exactly the true value falls in between the two distorted values p_+^0 and p_-^0 .

The literature has proposed alternative ways to determine p^0 . For example, it is common in experimental economics and survey design to randomize features that may affect responses but are deemed irrelevant. In our context, this means that we obtain

$$p_{AV}^0 = \frac{p_-^0 + p_+^0}{2}. \quad (10)$$

A more sophisticated method was proposed by Goldin and Reck (2019, 2020). They assume that some agents are consistent and always choose in line with their true preferences, while other agents are inconsistent and always follow the frame. Choices against a frame then identify all consistent agents with the respective opposing preference, so that

$$p_{GR}^0 = \frac{p_+^0}{p_+^0 + p_-^1} \quad (11)$$

identifies the share among the consistent agents who truly prefer $i = 0$. Under the additional assumption that consistent and inconsistent agents have the same distribution of preferences, (11) also identifies the value p^0 in the entire population.¹³

We now proceed to compare the three predictions (9)-(11) empirically. We use the data of Serra-Garcia and Szech (2023) on the impact of defaults on the willingness to vaccinate in the context of the COVID-19 pandemic. In the first wave of their experiment, Serra-Garcia and Szech (2023) recruited about 600 subjects on Prolific and elicited their intention to get vaccinated. Subjects were randomly assigned to one of three treatments: opt-out, opt-in, and active-choice. The opt-out treatment described a situation where an appointment for vaccination already exists, and the “Leave as is and receive the vaccine” answer option was preset as the default on the computer screen. Analogously, the opt-in treatment described a

¹³Goldin and Reck (2019, 2020) suggest the assumption of identical preferences only conditional on observable covariates. Since we do not use covariates to improve the other methods, we do not use them here either.

situation with no appointment as the preset default. The active-choice treatment described the problem neutrally without any preset default. Response times were recorded.¹⁴ While a neutral framing is often difficult or impossible to achieve in real-world applications, the existence of the neutral active-choice treatment in the experiment allows us to evaluate the predictions of all three methods against a ground truth.

The predictions are, despite being in principle designed for a within-subjects application, valid for the experimental between-subjects design under the assumption that random allocation of subjects to treatments generates samples representative of the population. For our response time-based prediction, we normalize response times by subtracting in logs a baseline response time, to account for individual heterogeneity (see Section 2.5.3). As there exists no appropriate binary baseline question in the experiment, we use the total duration it took subjects to complete the study net of the response time in the question about vaccination.

In Figure 3, the blue bars show the shares of subjects who decide to remain unvaccinated under opt-in ($p_-^0 = 0.342$), opt-out ($p_+^0 = 0.303$), and active-choice ($p^0 = 0.335$). The orange bars are our predictions for the active-choice outcome based on response time ($p_{RT}^0 = 0.338$), simple averaging ($p_{AV}^0 = 0.323$), and Goldin-Reck ($p_{GR}^0 = 0.315$). The response time-based method clearly outperforms both alternatives and gets very close to the actual choice share in the active-choice treatment.

We can also ask which of the two defaults induces the optimal choice for a larger number of subjects. Among the subjects who respond to the frame, a fraction $(p^0 - p_+^0)/(p_-^0 - p_+^0) = 0.821$ truly prefer to remain unvaccinated, and hence opt-in is optimal for a larger number of responsive subjects.¹⁵ If we replace the true value p^0 in this expression by the estimated value p_{RT}^0 , then we arrive at the same conclusion and thus correctly identify the optimal nudge. By contrast, using p_{AV}^0 by definition always implies that both frames are equally good, while using p_{GR}^0 here suggests opt-out as optimal.

To get an impression of how variable the three predictions are, we sample with replacement from the opt-in and opt-out treatment groups to generate synthetic treatment groups of the same size as the original data. We do this 5'000 times and each time compute the

¹⁴The data for all applications in Sections 3.2–3.4 were collected using Qualtrics. In Sections 3.3 and 3.4, we follow Liu and Netzer (2023) and define response time by the time of the last click before submission of the answer. We have to deviate from this rule here in Section 3.2, because the existence of a preset default answer made it possible to submit that answer without any earlier clicks, so that the time of the last click is often missing. We therefore use the time of submission of the answer here. Remember that it is not a problem for our approach if choosing the default is systematically faster as it requires fewer clicks than choosing against the default, as discussed in footnote 12.

¹⁵That statement about the optimal frame is subject to various qualifications. For example, it ignores possible welfare effects of the default on subjects who do not respond to changes in the default. It also ignores other welfare effects of vaccinations due to externalities etc.

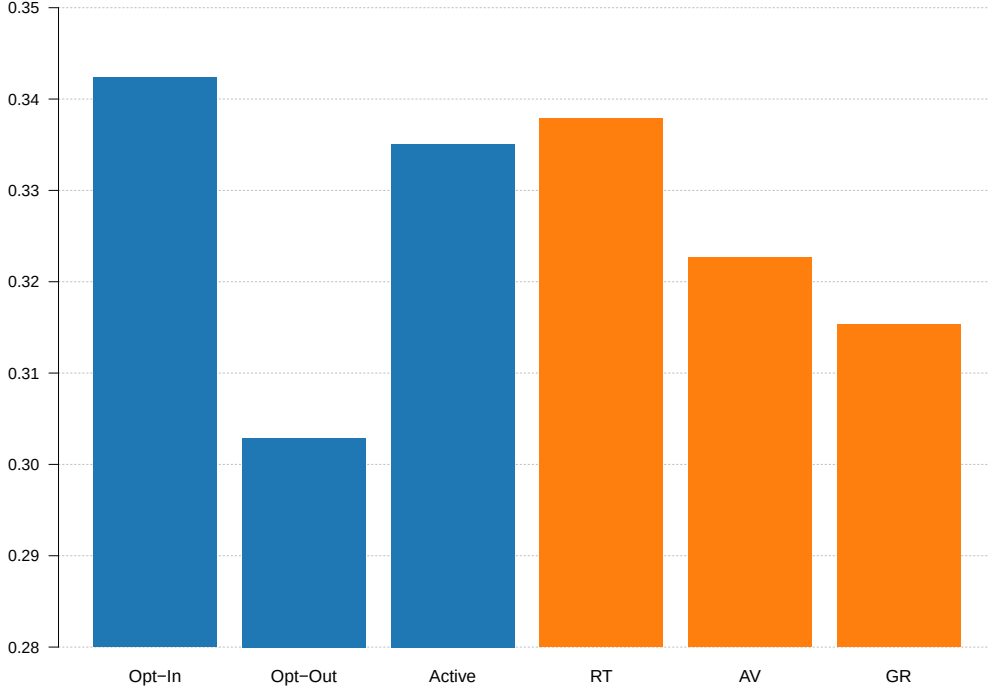


Figure 3: Shares unvaccinated

Notes: The blue bars depict the shares of subjects who decide to remain unvaccinated under the treatments opt-in, opt-out, and active-choice in the experiment of Serra-Garcia and Szech (2023). The orange bars depict the predictions of the share of unvaccinated subjects under active-choice for the three methods based on response times (RT), simple averaging (AV), and Goldin-Reck (GR).

predictions and their difference to the active-choice benchmark, which we keep fixed.¹⁶ The response time-based prediction remains the most precise estimate in 55.1% of the samples, followed by averaging in 22.8% and Goldin-Reck in 22.1% of all samples. The optimal nudge is correctly identified by the response time-based method in 68.7% of all samples, followed by averaging in 50.0% (by definition) and Goldin-Reck in 28.2% of all samples.

Overall, this application shows that response times can recover meaningful preference information even when choices are distorted by a frame. Using response times to identify the indifferent individual, our method yields a reliable estimate of the non-distorted choice share and identifies which frame best aligns with underlying preferences. The empirical comparison

¹⁶If the experiment were a within-subjects design, it would be appropriate to sample subjects and then use their choices across all three treatments in the synthetic samples. This is not feasible with the current between-subjects design, and hence we keep the active-choice benchmark as the fixed ground truth. When sampling from the original data, it can happen that the fraction of unvaccinated subjects becomes smaller under opt-in than under opt-out, effectively corresponding to a directional reversal of the standard default effect. We let the data speak in those cases and treat opt-in as the frame favoring vaccination.

highlights that this approach can outperform existing alternatives, both in accuracy and in its ability to guide the choice of an optimal nudge.

3.3 Testing Decreasing Marginal Happiness

Easterlin (2005) postulates that “[f]ew generalizations in the social sciences enjoy such wide-ranging support as that of diminishing marginal utility of income”, where he interprets utility explicitly as “subjective well-being” (p. 243). Yet the evidence is surprisingly fragile. Easterlin himself showed that the principle does not extend from cross-section to time series data. Later studies, including Oswald (2008) and Kaiser and Oswald (2022), emphasized that even cross-sectional patterns are inconclusive. The difficulty is that the observed relationship between income and reported happiness is the composition of two functions: the true mapping $h : W \rightarrow H$ from income to happiness and the reporting function $r : H \rightarrow R$ from happiness to survey response. Observing that $r(h(w))$ is concave does not imply that $h(w)$ is concave and therefore does not establish decreasing marginal happiness. The relationship may just as well be due to a concave reporting function $r(h)$. Other studies have estimated advanced ordered response models, but as Bond and Lang (2019) have shown, their conclusions depend on the distributional assumptions of the model. Kaiser and Oswald (2022) conclude that the problem is “fundamental, little recognized, and so far unsolved” (p. 3).

To formalize the income-happiness relation in our framework, let $(G_w)_w$ denote the family of happiness distributions for all possible income levels $w \in W \subseteq \mathbb{R}_+$ and define $\mu(w) = \int_{\mathbb{R}} x dG_w(x)$ as the average happiness of agents with income w . The question of interest is whether μ is a concave function of w . A natural starting point is to plot the average response in a happiness survey against income, as many studies have done. We repeat this exercise here, using MTurk survey data of Liu and Netzer (2023) in which 3’743 respondents answered a binary question on whether they are “rather happy” or “rather unhappy.” Household income in the survey was reported in three broad bins: below \$40’000, between \$40’000 and \$69’999, and above \$70’000. To proceed, we need to associate each of these bins with a unique income value, denoted w_L , w_M and w_H . As a first illustrative benchmark, we take the midpoint of each bin, assuming an upper bound of \$110’000 for the high-income bin. Figure 4 then shows that the share of “rather happy” respondents rises with income but at a diminishing rate, producing a concave pattern. Like in Easterlin (2005) and the literature cited therein, this is a bivariate relation without any additional controls, but the results are similar when conditioning on various socio-demographic variables.

Our point is that concavity in Figure 4 does not establish concavity of μ , which here

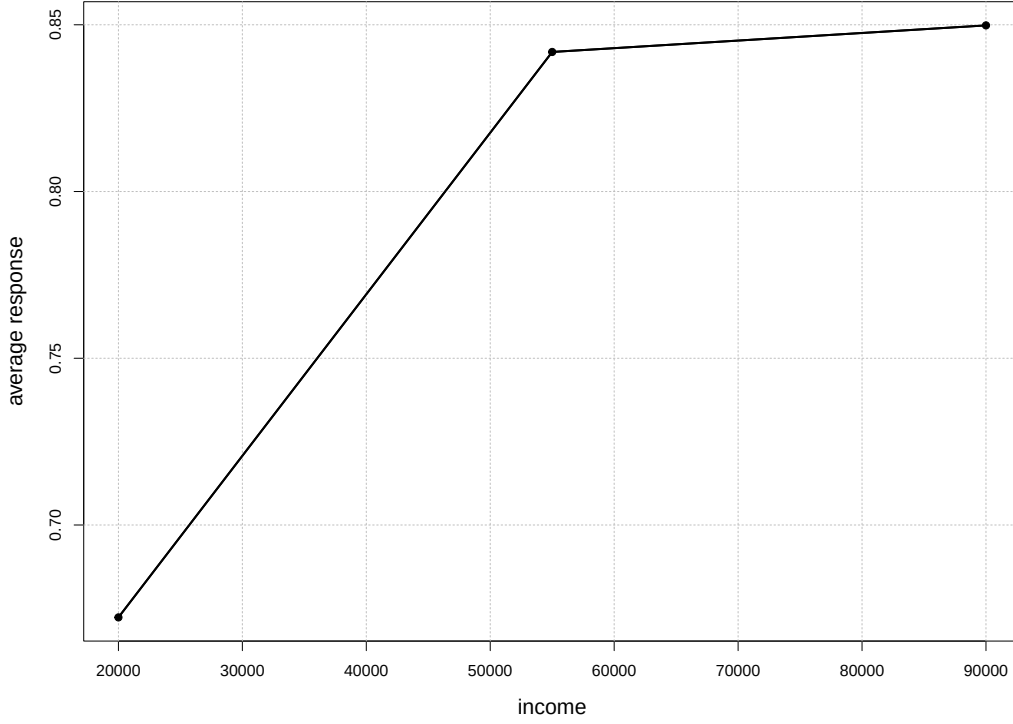


Figure 4: The relationship between income and reported happiness.

Notes: The curve corresponds to the case where the income for each bin is given by the midpoint of the bin, assuming an upper bound of \$110'000 for the high-income bin. Responses are coded as 0 (“rather unhappy”) and 1 (“rather happy”), so the average coincides with the share of respondents reporting “rather happy.”

reduces to the single inequality

$$\alpha\mu(w_L) + (1 - \alpha)\mu(w_H) \leq \mu(w_M), \quad (12)$$

where α is such that $\alpha w_L + (1 - \alpha)w_H = w_M$ (and which becomes $\alpha = 0.5$ in our illustrative benchmark). Following the logic of Bond and Lang (2019), we can explain Figure 4 with happiness distributions (G_L, G_M, G_H) for which (12) is violated, for example distributions that become more right-skewed as income grows and which therefore go along with high average happiness among the rich. It is impossible to reject or verify such distributional assumptions based on response data alone.

According to (12), decreasing marginal happiness is equivalent to a ranking of the means of two distributions. The LHS of (12) represents the mean happiness of a population that mixes low- and high-income individuals in proportions α and $1 - \alpha$, while the RHS is the

mean happiness of the middle-income group. We will now develop response time-based criteria that allow us to rank these means.

The ranking of the means of two distributions G_A and G_B is invariant to positive affine transformations that are identical for the two groups. Detecting it thus requires knowledge of the chronometric functions up to common and positive linear transformations. Since this may be too demanding, we can again work with a sufficient condition. Consider the asymmetry property that $G_A(x) + G_A(-x) \leq G_B(x) + G_B(-x)$ for all $x \in \mathbb{R}_+$, which, as shown in Appendix C, implies that the mean of G_A is larger than the mean of G_B . This condition is invariant to transformations that are identical across groups and symmetric around zero. Therefore, an assumption guaranteeing that it can be detected from response time data is that the chronometric effect is identical for the two options and the two groups, but otherwise unrestricted. Theorem 1 then implies that the condition is detected if it holds for functions H_A and H_B constructed based on (5). This can once more be expressed directly in terms of the observed data, and it follows that

$$p_A^0 F_A^0(t) - p_B^0 F_B^0(t) \leq p_A^1 F_A^1(t) - p_B^1 F_B^1(t) \quad (13)$$

for all $t \in [\underline{t}, \bar{t}]$ is sufficient for detecting that the mean of G_A is larger than the mean of G_B . An analogous statement holds for detecting a strict inequality of the means.¹⁷

A stronger sufficient condition for the ranking of two means is first-order stochastic dominance, $G_A(x) \leq G_B(x)$ for all $x \in \mathbb{R}$. FOSD is attractive because it is invariant to any pair of identical increasing transformations. Detecting FOSD therefore requires only that the chronometric effect is identical for both groups; since no comparison between positive and negative values is involved, we do not have to assume that the chronometric function is the same for both options. Proceeding as before then yields the stronger condition

$$p_A^0 F_A^0(t) - p_B^0 F_B^0(t) \leq 0 \leq p_A^1 F_A^1(t) - p_B^1 F_B^1(t) \quad (14)$$

for all $t \in [\underline{t}, \bar{t}]$ to detect FOSD. An analogous statement holds when strict inequality for some $x \in \mathbb{R}$ is required in the definition of FOSD.

Condition (14) equals conditions (i) and (ii) of Proposition 2 in Liu and Netzer (2023).¹⁸ They test these inequalities in survey data and show that the null hypothesis of FOSD often

¹⁷Condition (13) was first derived in our earlier working paper Liu and Netzer (2020). We referenced it in a footnote of Liu and Netzer (2023) but have not published it otherwise.

¹⁸Proposition 2 in Liu and Netzer (2023) characterizes detection of first-order stochastic dominance using general ordered response models for surveys with more than two response categories. Their condition (iii) applies to the intermediate response categories and coincides with the demanding condition by Bond and Lang (2019), as response times are not monotone and hence not informative in the intermediate categories.

cannot be rejected, in particular in cases where a probit model yields significant parameter estimates, indicating that the results of conventional models often seem to be qualitatively robust. Castillo and Vitaku (2025) have subsequently used the condition to confirm the robustness of their test of salience theory, and Cheng et al. (2025) have used it for evaluating the causal impact of bullet-chatting on online viewer behavior.

We now use these conditions to test the hypothesis that marginal happiness is decreasing in income, as formalized in (12).¹⁹ Under the assumption that all three income groups have the same chronometric function, which our criteria require anyway, we mix the two extreme income groups with appropriate weights to obtain the cdf $G_P(x) = \alpha G_L(x) + (1 - \alpha)G_H(x)$ of happiness in the pooled group. Condition (13) then yields

$$p_M^0 F_M^0(t) - p_P^0 F_P^0(t) \leq p_M^1 F_M^1(t) - p_P^1 F_P^1(t) \quad (15)$$

for all $t \in [\underline{t}, \bar{t}]$, with $p_P^i F_P^i(t) = \alpha p_L^i F_L^i(t) + (1 - \alpha)p_H^i F_H^i(t)$, as a sufficient condition to detect decreasing marginal happiness. Intuitively, this inequality rules out examples like the increasingly right-skewed distributions discussed above, as these would generate relatively fast happy and slow unhappy responses in the pooled group.

We follow Liu and Netzer (2023) and normalize response times by subtracting in logs the response time to a baseline marital-status question. This normalization corroborates further the assumption of identical chronometric functions in the different income groups.²⁰ Figure 5 plots the two sides of (15) for the illustrative benchmark of $\alpha = 0.5$. The inequality is not exactly satisfied in the data, because the LHS exceeds the RHS for some small t . To assess statistical significance, we employ a bootstrap-based method as in Liu and Netzer (2023) that rests on a test by Barrett and Donald (2003). To ensure robustness, we also go beyond

¹⁹There are similar yet different approaches in the literature. Already in his 1996 opus magnum *Infinite Jest*, novelist David Foster Wallace described a related idea. The novel centers around a lost movie that is so entertaining that people who have seen it are willing to sacrifice everything, including their toes, to see it again. A terrorist organization wants to use the movie as a weapon and contemplates how to test that the marginal entertainment of the movie is not decreasing. They propose checking that the time it takes for people to accept sacrificing another toe is not increasing after repeated views of the movie. We remark that this procedure does not cleanly distinguish between the marginal entertainment of the movie and the marginal cost of losing a toe. A non-fiction investigation of how response times vary with value levels in an individual decision-making task was done by Shevlin et al. (2022). They show that response times tend to decrease in the value level, incompatible with decreasing marginal utility. Liu and Netzer (2023) have shown that, using an ordered probit model, happiness increases significantly when moving from the low to the middle income category and that, using response times, the hypothesis of FOSD of the happiness distributions cannot be rejected. The effect of moving from middle to high income is insignificant and FOSD can be rejected. These results do not yet establish concavity of the income-happiness relation.

²⁰Our results in Subsection 2.5.3 imply that all properties which we investigate here can be tested using the response time-based detection conditions: under the null hypothesis that the property holds, its detection condition must be satisfied in the normalized data even when there is noise or measurement error.

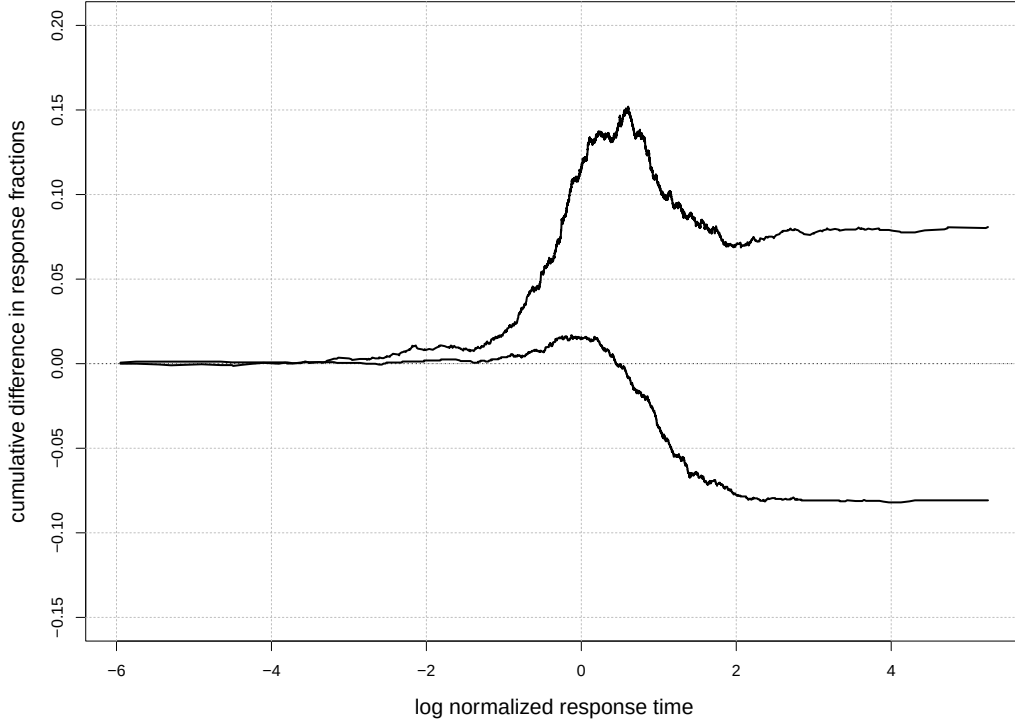


Figure 5: Empirical conditions for testing the income-happiness relation.

Notes: The curve (depicted for the case $\alpha = 0.5$) at the top represents the empirical function $p_M^1 F_M^1(t) - p_P^1 F_P^1(t)$, while the one at the bottom represents $p_M^0 F_M^0(t) - p_P^0 F_P^0(t)$.

the benchmark $\alpha = 0.5$ and conduct the test separately for each value $\alpha \in \{0.1, 0.2, \dots, 0.9\}$, i.e., we vary the mixing weight used to construct the pooled group P .²¹ As Table 1 shows, the p-values for the null hypothesis that condition (15) holds are consistently large. We clearly cannot reject our sufficient condition for decreasing marginal happiness.

We also test the stronger property that G_M first-order stochastically dominates G_P , which is a fortiori sufficient for decreasing marginal happiness but whose detection does not require symmetry of the chronometric functions. The testing condition here becomes

$$p_M^0 F_M^0(t) - p_P^0 F_P^0(t) \leq 0 \leq p_M^1 F_M^1(t) - p_P^1 F_P^1(t) \quad (16)$$

²¹Different choices of representative incomes for the survey bins imply different values of α through the identity $\alpha w_L + (1 - \alpha)w_H = w_M$. For instance, changing the upper bound for the high-income group from \$110'000 to \$150'000 or \$200'000 results in $\alpha = 0.611$ or $\alpha = 0.696$, respectively. Alternatively, one could use more granular income data about MTurk subjects drawn from other studies to approximate the average income in each bin. Doing this with the data from Moss et al. (2023, Table 2), and using an upper bound of \$200'000 for the high-income bin, we obtain $\alpha = 0.646$.

α	p-value	
	(15)	(16)
0.1	0.962	0.195
0.2	0.961	0.205
0.3	0.958	0.251
0.4	0.956	0.298
0.5	0.955	0.351
0.6	0.949	0.405
0.7	0.941	0.449
0.8	0.941	0.517
0.9	0.934	0.566

Table 1: Summary of tests for concavity.

for all $t \in [\underline{t}, \bar{t}]$, which in Figure 5 requires that the functions are separated by zero. While not exactly satisfied in the data, Table 1 shows that the p-values remain large.²² We clearly cannot reject even this stronger sufficient condition for decreasing marginal happiness.

Finally, we investigate the hypothesis that marginal happiness increases with income, which corresponds to reversing all inequalities studied so far. The reversed versions of (15) and (16) are rejected at all plausible levels of statistical significance (all $p = 0.000$).

To summarize, our results are consistent with the idea that marginal happiness is decreasing in income in a cross-sectional data set. While our tests avoid several of the pitfalls noted earlier in this section, some limitations remain. Most notably, we were only able to test the concavity of expected happiness for three income levels implied by the survey bins, rather than across the entire income distribution. Additionally, the income-happiness relation is bivariate without controlling for potential confounding factors. Future research that addresses these limitations holds great promise.

3.4 Predicting Treatment Heterogeneity

Policymakers and experimenters frequently need to decide which populations to target with a treatment. For instance, a policymaker may want to make costly employment training available only to the most responsive subgroup of the unemployed. An online platform may want to increase prices for those consumers who are least likely to reduce demand. A tax authority following the Ramsey (1927) rule wants to impose larger taxes on markets with

²²The test implements the procedure based on Barrett and Donald (2003) together with a joint hypothesis correction by Romano and Wolf (2016) to account for the fact that condition (16) contains two inequalities. We refer the reader to Liu and Netzer (2023) for details.

smaller elasticities. An experimenter may want to impose some treatment on the least and the most responsive group to obtain bounds on the treatment effect.

Modern methods for heterogeneous treatment effects and policy learning, such as causal forests or empirical welfare learning, rely on post-treatment outcomes to estimate subgroup effects (e.g., Athey and Imbens, 2016; Kitagawa and Tetenov, 2018). Likewise, behavioral work on treatment moderators relies on comparisons at the post-treatment level (Krefeld-Schwalb, Sugerman and Johnson, 2024). In other words, the treatment must be imposed on a sample of the entire population before the eventual target subgroups can be identified. We show that the ordering of treatment effects across subgroups can be predicted pre-treatment, using choices and response times from a baseline before any treatment even took place. As such, our approach may be used as a pre-outcome data source when developing statistical treatment rules under limited information as in Manski (2004).

Let G_A and G_B denote the latent distributions for groups A and B , respectively. Before treatment, the probability that an agent in group j chooses $i = 1$ is $1 - G_j(0)$. Assume that the agents then receive a treatment that favors option $i = 1$. We model this by adding a shock $\delta \geq 0$ to each agent's latent value. After treatment, the probability of choosing $i = 1$ is $1 - G_j(-\delta)$, so that the treatment effect for group j is $G_j(0) - G_j(-\delta) \geq 0$. This treatment effect is larger for group A than for group B whenever

$$G_B(0) - G_B(-\delta) \leq G_A(0) - G_A(-\delta). \quad (17)$$

If we do not know the exact size of δ , we can still predict that group A will react more strongly than group B when we detect that (17) holds for all $\delta \geq 0$. This property is invariant to transformations (ψ_A, ψ_B) which satisfy $\psi_A(x) = \psi_B(x)$ for all $x \leq 0$ and $\psi_A(0) = \psi_B(0) = 0$. A natural assumption guaranteeing that the chronometric functions are known up to these transformations is that the chronometric function when choosing $i = 0$ is identical for the two groups. Then, under this assumption, the distributions H_j constructed by (5) satisfy the desired inequality if and only if

$$p_B^0(1 - F_B^0(t)) \leq p_A^0(1 - F_A^0(t)) \text{ for all } t \in [t, \bar{t}]. \quad (18)$$

This condition rests on data from choices of $i = 0$ only, as these are the ones that the treatment may potentially shift. The condition requires $p_B^0 \leq p_A^0$ to be satisfied and, additionally, that the choices of group A must be sufficiently slow compared to group B . Intuitively, considering the agents who take at least time t to respond picks out those who are relatively close to indifferent, at the resolution defined by t . If group A has more of this near-indifferent mass than group B for any t , then a treatment that shifts the latent index to the right con-

verts more mass into $i = 1$ choices in group A than in group B .²³ Taken together, (18) is a sufficient condition that allows us to rank groups by responsiveness using choices and response times pre-treatment and no further distributional assumptions.

If an analyst has more information about the size of the treatment shock δ , the detection condition can be weakened. For example, the analyst may have prior knowledge that the treatment effect is bounded such that it never shifts the decision of the $q\%$ most committed supporters of option $i = 0$ among the entire population. Inequality (18) then needs to be satisfied only for all $t \in [t_q, \bar{t}]$, where t_q is the respective percentile of the population response time distribution and given by $F^0(t_q) = q/100$. Thus, the bound on the treatment effect translates into a natural weaker condition: we only need to compare the near-indifferent masses at sufficiently slow response times. Consequently, and in contrast to the unbounded case, the detection condition does not necessarily require $p_B^0 \leq p_A^0$, that is, we could have a larger number of agents favoring $i = 0$ in group B than in group A and still predict a stronger treatment response from group A .

We put these results to work using data from Krefeld-Schwalb, Sugerman and Johnson (2024). In their Study 3, a total of 1'200 subjects recruited on Prolific and MTurk responded to the Trolley Problem. In this moral dilemma, a runaway trolley threatens the lives of five people. In the *track case*, which the authors take as the baseline, pulling a lever diverts the trolley onto a side track, killing one instead. In the *bridge case*, which serves as the treatment, the only way to stop the trolley is by pushing a man onto the tracks. While the two scenarios are identical in that an intervention results in only one death instead of five, they are typically judged differently by subjects. In the data of Krefeld-Schwalb, Sugerman and Johnson (2024), 83.6% of all subjects report that pulling the lever in the track case is morally admissible, while only 43.3% of all subjects report that pushing the man onto the tracks in the bridge case is morally admissible. Krefeld-Schwalb, Sugerman and Johnson (2024) are interested in treatment heterogeneity at the subject level. They develop a factor model encompassing the effect moderators Fluid Intelligence (F), Attentiveness (A), Crystallized Intelligence (C), and Experience (E). They find that attention is the only moderator that has a statistically significant influence on the treatment effect, with more attentive subjects reacting more strongly.

Our goal is to use subjects' responses and response times in the track case to predict

²³The literature has documented several times that agents who respond more slowly are more likely to change their behavior over time (e.g. Fazio and Williams, 1986; Bassili, 1993; Kononov and Krajchich, 2019), across elicitation modes (e.g. Alós-Ferrer et al., 2016), or in response to counterarguments (e.g. Bassili and Fletcher, 1991; Huckfeldt et al., 1999). However, there are also experiments in which response time does not systematically predict behavioral change (Bassili and Krosnick, 2000). Our contribution here is to derive a precise criterion for predicting treatment heterogeneity and to test this criterion empirically.

how the treatment effect of moving to the bridge case differs across subgroups. In terms of our above model, option $i = 0$ corresponds to the response that the intervention is morally admissible, which goes down with the treatment. We construct pairs of groups using the demographic information available in the study as well as the four factors that Krefeld-Schwalb, Sugerman and Johnson (2024) consider as moderators. To predict which of two groups has the stronger treatment effect, we test condition (18) in both directions. If only one of these hypotheses is rejected, we predict that the treatment effect is larger for the respective group. If both or none of the hypotheses are rejected, we do not make a prediction.²⁴ We then check the correctness of our prediction using the actual data from the bridge case.

To form subgroups, we split the sample at the median along the respective dimension and compare the top and bottom halves with each other. For gender, we compare women and men. Table 2 reports p-values, predictions and actual outcomes for all pairs of groups. In 7 out of 9 cases, we are able to make a prediction at the 5% significance level. All of these predictions turn out to be correct. In particular, it is instructive to compare our predictions for the FACE moderators to Krefeld-Schwalb, Sugerman and Johnson (2024). They find that more attentive subjects exhibit a significantly stronger treatment effect, while not documenting any significant effects for the other three moderators. Our analysis, based on pre-treatment data only, reveals the same prediction for attention. Moreover, our (significant) predictions for the other three moderators broadly agree with the (insignificant) signs reported in Krefeld-Schwalb, Sugerman and Johnson (2024, Table S17).²⁵

Our approach does not yield predictions when we consider subgroup splits based on education and income, because neither direction of inequality (18) can be rejected. A possible way forward is to assume that the treatment effect is bounded and to exclude the most committed (fastest) subjects from the test. However, even in that case we do not obtain predictions at the 5% confidence level.

Overall, our analysis shows that response times can be used to anticipate heterogeneous treatment effects before the intervention takes place. By translating our detection condition (18) into a testable inequality, we can successfully rank groups by anticipated responsiveness using only baseline choices and response times.

²⁴The hypothesis test relies on a bootstrapping procedure as described in Section 3.3. Our results in Subsection 2.5.3 ensure the validity of our tests even when there is noise or measurement error. We again log-normalize response times. Since no appropriate binary question is available, we proceed as in Section 3.2 and use the total duration it took subjects to complete the study net of the focal question’s response time.

²⁵More precisely, in their specification with only the four moderators, the signs agree for all four moderators. In their specification with the moderators and a dummy for the panel (Prolific/MTurk) the signs agree for attention, crystallized intelligence, and fluid intelligence, but not for experience.

Dimension (group A vs. B)	$p(A > B)$	$p(B > A)$	Prediction	Correct
Gender (women vs men)	0.872	0.005	A stronger	True
Politics (liberal vs conservative)	0.790	0.000	A stronger	True
Income (poor vs rich)	0.477	0.510	-	-
Education (low vs high)	0.065	0.078	-	-
Age (young vs old)	0.000	0.569	B stronger	True
Face F (top vs bottom)	0.583	0.000	A stronger	True
Face A (top vs bottom)	0.896	0.000	A stronger	True
Face C (top vs bottom)	0.001	0.590	B stronger	True
Face E (top vs bottom)	0.000	0.387	B stronger	True

Table 2: Testing the ranking condition in (18) in both directions.

Notes: The table reports p-values from testing the sufficient pre-treatment ranking condition in (18) in both directions for each pair of subgroups (group A vs. B). We predict which subgroup exhibits the stronger treatment effect of moving to the bridge case when exactly one direction is rejected at the 5% level; otherwise no prediction is made. Subgroups are formed by median splits for demographics and the FACE moderators (top vs. bottom halves).

4 Conclusion

The goal of this paper is to provide a systematic account of the information that response time data contain. We approach the problem by phrasing it as one of identification in the context of binary response models. Our main theoretical result relates the set of identifiable distributional properties to the set of admissible chronometric functions. The fundamental idea is that the joint distribution of responses and response times identifies a composition of the latent distribution and the chronometric function. Properties of the distribution that are preserved under a given set of transformations can therefore be identified if the chronometric function is known up to these transformations. Several existing results in the literature follow as corollaries and can be generalized. Many new results emerge. We demonstrate the usefulness of our approach in four empirical applications: revealed preference analysis, identifying an optimal nudge, testing decreasing marginal happiness of income, and predicting treatment heterogeneity. In each case, incorporating response times yields new economic insights that cannot be obtained from conventional methods alone.

Our applications in Section 3 are merely examples of the scope of the method and not an exhaustive list. Appendix E presents additional distributional properties that can be studied within our framework, including the ranking of inequality or dispersion, likelihood-ratio dominance, unimodality, and correlations between latent variables and observable charac-

teristics. These properties arise naturally in many empirical settings. Applications that we have in mind include the study of polarization using surveys on political attitudes, evaluating the magnitude of under-reporting on sensitive views based on behavioral records in the lab, and optimal product pricing using data on purchase decisions from online platforms. There are also several possible extensions of our framework that merit further investigation. These include correlations among multiple latent variables, the case with more than two choice options, and the use of response times when they are affected by additional factors like player types or decision modes as in Rubinstein (2007, 2013, 2016).

References

- Alós-Ferrer, Carlos, and Michele Garagnani.** 2022. “Strength of Preference and Decisions Under Risk.” *Journal of Risk and Uncertainty*, 64: 309–329.
- Alós-Ferrer, Carlos, and Michele Garagnani.** 2024. “Improving Risky-Choice Predictions Using Response Times.” *Journal of Political Economy: Microeconomics*, 2(2): 335–354.
- Alós-Ferrer, Carlos, Dura-Georg Granić, Johannes Kern, and Alexander K. Wagner.** 2016. “Preference Reversals: Time and Again.” *Journal of Risk and Uncertainty*, 52: 65–97.
- Alós-Ferrer, Carlos, Ernst Fehr, and Nick Netzer.** 2021. “Time Will Tell: Recovering Preferences When Choices Are Noisy.” *Journal of Political Economy*, 129(6): 1828–1877.
- Alós-Ferrer, Carlos, Ernst Fehr, Helga Fehr-Duda, and Michele Garagnani.** 2024. “Common Ratio and Common Consequence Effects Arise From True Preferences.” Mimeo.
- Alós-Ferrer, Carlos, Michele Garagnani, and Ernst Fehr.** 2024. “Identifying Non-transitive Preferences.” Mimeo.
- Arrow, Kenneth J.** 1958. “Rational Choice Functions and Orderings.” *Economica*, 26(102): 121–127.
- Athey, Susan.** 2002. “Monotone Comparative Statics Under Uncertainty.” *The Quarterly Journal of Economics*, 117(1): 187–223.
- Athey, Susan, and Guido W. Imbens.** 2016. “Recursive Partitioning for Heterogeneous Causal Effects.” *Proceedings of the National Academy of Sciences*, 113(27): 7353–7360.

- Atkinson, Anthony B.** 1970. “On the Measurement of Inequality.” *Journal of Economic Theory*, 2(3): 244–263.
- Barrett, Garry F., and Stephen G. Donald.** 2003. “Consistent Tests for Stochastic Dominance.” *Econometrica*, 71(1): 71–104.
- Bassili, John N.** 1993. “Response Latency Versus Certainty as Indexes of the Strength of Voting Intentions in a CATI Survey.” *Public Opinion Quarterly*, 57(1): 54–61.
- Bassili, John N., and Jon A. Krosnick.** 2000. “Do Strength-Related Attitude Properties Determine Susceptibility to Response Effects? New Evidence From Response Latency, Attitude Extremity, and Aggregate Indices.” *Political Psychology*, 21(1): 107–132.
- Bassili, John N., and Joseph F. Fletcher.** 1991. “Response-Time Measurement in Survey Research: A Method for Cati and a New Look at Nonattitudes.” *Public Opinion Quarterly*, 55(3): 331–346.
- Benkert, Jean-Michel, and Nick Netzer.** 2018. “Informational Requirements of Nudging.” *Journal of Political Economy*, 126(6): 2323–2355.
- Bond, Timothy N., and Kevin Lang.** 2019. “The Sad Truth About Happiness Scales.” *Journal of Political Economy*, 127(4): 1629–1640.
- Bucher, T., C. Collins, M.E. Rollo, T.A. McCaffrey, N. De Vlieger, D. Van der Bend, H. Truby, and F.J.A. Perez-Cueto.** 2016. “Nudging Consumers Towards Healthier Choices: A Systematic Review of Positional Influences on Food Choice.” *British Journal of Nutrition*, 115: 2252–2263.
- Card, David, Stefano DellaVigna, Chenxi Jiang, and Dmitry Taubinsky.** 2024. “Understanding Expert Choices Using Decision Time.” Mimeo.
- Castillo, Marco, and Valon Vitaku.** 2025. “Salience in Choice Under Risk: An Experimental Investigation.” Mimeo.
- Chabris, Christopher F., Carrie L. Morris, Dmitry Taubinsky, David Laibson, and Jonathon P. Schuldt.** 2009. “The Allocation of Time in Decision-Making.” *Journal of the European Economic Association*, 7(2-3): 628–637.
- Cheng, Yan, Jingbo Wang, Xinyu Cao, and Zuo-Jun Max Shen.** 2025. “Do Bullet Chats Keep Viewers Watching? Estimating Heterogeneous Treatment Effects with a Control Function Approach.” Mimeo.

- Clithero, John A.** 2018. “Improving Out-Of-Sample Predictions Using Response Times and a Model of the Decision Process.” *Journal of Economic Behavior and Organization*, 148: 344–375.
- Coffman, Katherine B, Lucas C Coffman, and Keith M Marzilli Ericson.** 2017. “The Size of the LGBT Population and the Magnitude of Antigay Sentiment Are Substantially Underestimated.” *Management Science*, 63(10): 3168–3186.
- Cotet, Miruna, Wenjia Joyce Zhao, and Ian Krajbich.** 2025. “Deliberation During Online Bargaining Reveals Strategic Information.” *Proceedings of the National Academy of Sciences*, 122(7): e2410956122.
- Cowell, Frank Alan.** 2011. *Measuring Inequality*. Oxford University Press.
- d’Aspremont, Claude, and Louis Gevers.** 2002. “Social Welfare Functionals and Interpersonal Comparability.” In *Handbook of Social Choice and Welfare*. Vol. 1, , ed. K.J. Arrow, A. K. Sen and K. Suzumura, Chapter 10, 459–541. Elsevier.
- Diamond, Peter, and Joseph Stiglitz.** 1974. “Increases in Risk and in Risk Aversion.” *Journal of Economic Theory*, 8(3): 337–360.
- DiMaggio, Paul, John Evans, and Bethany Bryson.** 1996. “Have American’s Social Attitudes Become More Polarized?” *American Journal of Sociology*, 102(3): 690–755.
- Dutta, Indranil, and James Foster.** 2013. “Inequality of Happiness in the US: 1972–2010.” *Review of Income and Wealth*, 59(3): 393–415.
- Easterlin, Richard A.** 2005. “Diminishing Marginal Utility of Income? Caveat Emptor.” *Social Indicators Research*, 70(3): 243–255.
- Echenique, Federico, and Kota Saito.** 2017. “Response Time and Utility.” *Journal of Economic Behavior & Organization*, 139: 49–59.
- Evans, John H.** 2003. “Have Americans’ Attitudes Become More Polarized? – An Update.” *Social Science Quarterly*, 84(1): 71–90.
- Fan, Yanqin, and Andrew J Patton.** 2014. “Copulas in Econometrics.” *Annual Review of Economics*, 6(1): 179–200.
- Fazio, Russell H.** 1990. “A Practical Guide to the Use of Response Latency in Social Psychological Research.” In *Research Methods in Personality and Social Psychology*. , ed. Clyde Hendrick and Margaret S. Clark, 74–97. Sage Publications.

- Fazio, Russell H., and Carol J. Williams.** 1986. "Attitude Accessibility as a Moderator of the Attitude-Perception and Attitude-Behavior Relations: An Investigation of the 1984 Presidential Election." *Journal of Personality and Social Psychology*, 51(3): 505–514.
- Ferrando, Pere J., and Urbano Lorenzo-Seva.** 2007*a*. "An Item Response Theory Model for Incorporating Response Time Data in Binary Personality Items." *Applied Psychological Measurement*, 31(6): 525–543.
- Ferrando, Pere J., and Urbano Lorenzo-Seva.** 2007*b*. "A Measurement Model for Likert Responses That Incorporates Response Time." *Multivariate Behavioral Research*, 42(4): 675–706.
- Fudenberg, Drew, Philipp Strack, and Tomasz Strzalecki.** 2018. "Speed, Accuracy, and the Optimal Timing of Choices." *American Economic Review*, 108(12): 3651–3684.
- Goldin, Jacob, and Daniel Reck.** 2019. "The Analysis of Survey Data with Framing Effects." *The American Statistician*, 73(3): 264–272.
- Goldin, Jacob, and Daniel Reck.** 2020. "Revealed-Preference Analysis with Framing Effects." *Journal of Political Economy*, 128(7): 2759–2795.
- Gonçalves, Duarte.** 2024. "Speed, Accuracy, and Complexity." Mimeo.
- Haile, Philip A., Ali Hortaçsu, and Grigory Kosenok.** 2008. "On the Empirical Content of Quantal Response Equilibrium." *American Economic Review*, 98(1): 180–200.
- Hammond, John S.** 1974. "Simplifying the Choice Between Uncertain Prospects Where Preference Is Nonlinear." *Management Science*, 20(7): 1047–1072.
- Haugh, Martin.** 2016. "An Introduction to Copulas." Lecture Notes, www.columbia.edu/~mh2078/QRM/Copulas.pdf.
- Hefti, Andreas, Shuo Liu, and Armin Schmutzler.** 2022. "Preferences, Confusion and Competition." *The Economic Journal*, 132(645): 1852–1881.
- Huckfeldt, Robert, Jeffrey Levine, William Morgan, and John Sprague.** 1999. "Accessibility and the Political Utility of Partisan and Ideological Orientations." *American Journal of Political Science*, 43(3): 888–911.
- Johnson, Justin P, and David P Myatt.** 2006. "On the Simple Economics of Advertising, Marketing, and Product Design." *American Economic Review*, 96(3): 756–784.

- Johnson, Martin.** 2004. “Timepieces: Components of Survey Question Response Latencies.” *Political Psychology*, 25(5): 679–702.
- Kaiser, Caspar, and Andrew J. Oswald.** 2022. “Inequality, Well-Being, and the Problem of the Unknown Reporting Function.” *Proceedings of the National Academy of Sciences*, 119(50): e2217750119.
- Kalmijn, Wim, and Ruut Veenhoven.** 2005. “Measuring Inequality of Happiness in Nations: In Search for Proper Statistics.” *Journal of Happiness Studies*, 6(4): 357–396.
- Kellogg, W. N.** 1931. “The Time of Judgment in Psychometric Measures.” *American Journal of Psychology*, 43(1): 65–86.
- Kendall, Maurice George.** 1955. *Rank Correlation Methods*. New York: Hafner Publishing Co.
- Kiefer, Nicholas M.** 1988. “Economic Duration Data and Hazard Functions.” *Journal of Economic Literature*, 26(2): 646–679.
- Kitagawa, Toru, and Aleksey Tetenov.** 2018. “Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice.” *Econometrica*, 86(2): 591–616.
- Konovalov, Arkady, and Ian Krajbich.** 2019. “Revealed Strength of Preference: Inference From Response Times.” *Judgment & Decision Making*, 14(4): 381–394.
- Krajbich, Ian, Bastiaan Oud, and Ernst Fehr.** 2014. “Benefits of Neuroeconomic Modeling: New Policy Interventions and Predictors of Preference.” *American Economic Review: Papers & Proceedings*, 105(5): 501–506.
- Krajbich, Ian, Dingchao Lu, Colin Camerer, and Antonio Rangel.** 2012. “The Attentional Drift-Diffusion Model Extends to Simple Purchasing Decisions.” *Frontiers in Psychology*, 3: 193.
- Krefeld-Schwalb, Antonia, Eli Rosen Sugerman, and Eric J. Johnson.** 2024. “Exposing Omitted Moderators: Explaining Why Effect Sizes Differ in the Social Sciences.” *Proceedings of the National Academy of Sciences*, 121(12): e2306281121.
- Lelkes, Yptach.** 2016. “Mass Polarization: Manifestations and Measurements.” *Public Opinion Quarterly*, 80(S1): 392–410.
- Liu, Shuo, and Nick Netzer.** 2020. “Happy Times: Identification From Ordered Response Data.” University of Zurich, Department of Economics, Working Paper No. 371.

- Liu, Shuo, and Nick Netzer.** 2023. "Happy Times: Measuring Happiness Using Response Times." *American Economic Review*, 113: 3289–3322.
- Luce, R. Duncan, David H. Krantz, Patrick Suppes, and Tversky Amos.** 1990. *Foundations of Measurement, Vol. III: Representation, Axiomatization, and Invariance*. Academic Press.
- Madrian, Brigitte C., and Dennis F. Shea.** 2001. "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior." *Quarterly Journal of Economics*, 116(4): 1149–1187.
- Manski, Charles F.** 1988. "Identification of Binary Response Models." *Journal of the American Statistical Association*, 83(403): 729–738.
- Manski, Charles F.** 2004. "Statistical Treatment Rules for Heterogeneous Populations." *Econometrica*, 72(4): 1221–1246.
- Maskin, Eric, and John Riley.** 2000. "Asymmetric Auctions." *The Review of Economic Studies*, 67(3): 413–438.
- Matzkin, Rosa L.** 1992. "Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and The Binary Choice Models." *Econometrica*, 60(2): 239–270.
- Milgrom, Paul R.** 1981. "Good News and Bad News: Representation Theorems and Applications." *The Bell Journal of Economics*, 12(2): 380–391.
- Moffatt, Peter G.** 2005. "Stochastic Choice and the Allocation of Cognitive Effort." *Experimental Economics*, 8(4): 369–388.
- Moss, Aaron J, Cheskie Rosenzweig, Jonathan Robinson, Shalom N Jaffe, and Leib Litman.** 2023. "Is It Ethical to Use Mechanical Turk for Behavioral Research? Relevant Data From a Representative Survey of Mturk Participants and Wages." *Behavior Research Methods*, 55(8): 4048–4067.
- Moyer, Robert S., and Richard H. Bayer.** 1976. "Mental Comparison and the Symbolic Distance Effect." *Cognitive Psychology*, 8(2): 228–246.
- Oswald, Andrew J.** 2008. "On the Curvature of the Reporting Function From Objective Reality to Subjective Feelings." *Economics Letters*, 100(3): 369–372.

- Palmer, John, Alexander C. Huk, and Michael N. Shadlen.** 2005. “The Effect of Stimulus Strength on the Speed and Accuracy of a Perceptual Decision.” *Journal of Vision*, 5: 376–404.
- Quah, John K.-H., and Bruno Strulovici.** 2012. “Aggregating the Single Crossing Property.” *Econometrica*, 80(5): 2333–2348.
- Ramsey, F.P.** 1927. “A Contribution to the Theory of Taxation.” *The Economic Journal*, 37(145): 47–61.
- Ranger, Jochen.** 2013. “Modeling Responses and Response Times in Personality Tests With Rating Scales.” *Psychological Test and Assessment Modeling*, 55(4): 361–382.
- Romano, Joseph P., and Michael Wolf.** 2016. “Efficient Computation of Adjusted p-Values for Resampling-Based Stepdown Multiple Testing.” *Statistics and Probability Letters*, 113: 38–40.
- Rubinstein, Ariel.** 2007. “Instinctive and Cognitive Reasoning: A Study of Response Times.” *The Economic Journal*, 117: 1243–1259.
- Rubinstein, Ariel.** 2013. “Response Time and Decision Making: An Experimental Study.” *Judgment and Decision Making*, 8(5): 540–551.
- Rubinstein, Ariel.** 2016. “A Typology of Players: Between Instinctive and Contemplative.” *The Quarterly Journal of Economics*, 131(2): 859–890.
- Samuelson, Paul A.** 1938. “A Note on the Pure Theory of Consumer’s Behaviour.” *Economica*, 5(17): 61–71.
- Serra-Garcia, Marta, and Nora Szech.** 2023. “Incentives and Defaults Can Increase COVID-19 Vaccine Intentions and Test Demand.” *Management Science*, 69(2): 1037–1049.
- Shaked, Moshe, and J George Shanthikumar.** 2007. *Stochastic Orders*. Springer.
- Shevlin, Blair R.K., Stephanie M. Smith, Jan Hausfeld, and Ian Krajbich.** 2022. “High-Value Decisions Are Fast and Accurate, Inconsistent With Diminishing Value Sensitivity.” *Proceedings of the National Academy of Sciences*, 119(6): e2101508119.
- Spearman, Charles.** 1904. “The Proof and Measurement of Association Between Two Things.” *The American Journal of Psychology*, 15(1): 72–101.
- Stevenson, Betsey, and Justin Wolfers.** 2008. “Happiness Inequality in the United States.” *The Journal of Legal Studies*, 37(S2): S33–S79.

- Takane, Yoshio, and Justine Sergent.** 1983. “Multidimensional Scaling Models for Reaction Times and Same-Different Judgements.” *Psychometrika*, 48: 393–423.
- Thaler, R.H., and C.R. Sunstein.** 2003. “Libertarian Paternalism.” *American Economic Review, Papers and Proceedings*, 93: 175–179.
- Thaler, R.H., and C.R. Sunstein.** 2008. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven: Yale University Press.
- Thistle, Paul D.** 1989. “Ranking Distributions with Generalized Lorenz Curves.” *Southern Economic Journal*, 56(1): 1–12.
- Tversky, Amos, and Daniel Kahneman.** 1981. “The Framing of Decisions and the Psychology of Choice.” *Science*, 211(4481): 453–458.
- Vaeth, Martin.** 2023. “Voter Learning, Unidimensional Ideology, and Polarization.” Mimeo.
- Wang, Tao, and Ehud Lehrer.** 2024. “Weighted Utility and Optimism/Pessimism: A Decision-Theoretic Foundation of Various Stochastic Dominance Orders.” *American Economic Journal: Microeconomics*, 16(1): 210–223.

Appendices

A Rationalizability

In this appendix, we study the question of *rationalizability* of the data: when $\mathcal{G}^*$ and $\mathcal{C}^*$ are both restricted sets of distributions and chronometric functions, respectively, does there exist $((G_j)_j, (c_j)_j) \in \mathcal{G}^* \times \mathcal{C}^*$ that induces the given data $(p_j, F_j)_j$?

We give an answer to this question for the case in which both $\mathcal{G}^*$ and $\mathcal{C}^*$ are generated using sets of transformations with properties like before. Formally, let $\mathcal{C}^*$ be the *maximal set* generated by $((c_j^*)_j, \Psi)$, that is,

$$\mathcal{C}^* = \{(c_j)_j \in \mathcal{C} \mid (c_j)_j = (c_j^* \circ \psi_j)_j \text{ for some } (\psi_j)_j \in \Psi\}.$$

In words, not only can every $(c_j)_j \in \mathcal{C}^*$ be derived from $(c_j^*)_j$ using a composition with some $(\psi_j)_j \in \Psi$, but the composition of $(c_j^*)_j$ with any $(\psi_j)_j \in \Psi$ yields an element of $\mathcal{C}^*$. This requires that $\psi_j(0) = 0$ holds for all transformations in Ψ . Analogously, let

$$\mathcal{G}^* = \{(G_j)_j \in \mathcal{G} \mid (G_j)_j = (G_j^* \circ \phi_j)_j \text{ for some } (\phi_j)_j \in \Phi\}$$

be the maximal set generated by $((G_j^*)_j, \Phi)$. For example, if $(G_j^*)_j$ are strictly increasing cdfs and Φ is the set of all profiles of strictly increasing transformations, then $\mathcal{G}^*$ becomes the set of all profiles of strictly increasing cdfs. If, in addition, $(G_j^*)_j$ are symmetric around their means, then combined with a suitably defined set of symmetric transformations we obtain a set of profiles of distributions that are also symmetric.

For the following result, denote by $\Phi \circ \Psi^{-1} = \{(\phi_j \circ \psi_j^{-1})_j \mid (\phi_j)_j \in \Phi \text{ and } (\psi_j)_j \in \Psi\}$ the set of all profiles of functions which are compositions of the elements of Φ and the inverse elements of Ψ . Any such function is bijective and strictly increasing, hence continuous.

Theorem 2. *Let $\mathcal{C}^*$ and $\mathcal{G}^*$ be the maximal sets generated by $((c_j^*)_j, \Psi)$ and $((G_j^*)_j, \Phi)$, respectively. Then, data $(p_j, F_j)_j$ is rationalizable if and only if there exists $(L_j)_j \in \Phi \circ \Psi^{-1}$ such that $(G_j^* \circ L_j)_j = (H_j)_j$.*

Proof. Only-if-statement. Suppose there exists $((G_j)_j, (c_j)_j) \in \mathcal{G}^* \times \mathcal{C}^*$ that induces $(p_j, F_j)_j$. As shown in the proof of Theorem 1, there exists $(\psi_j)_j \in \Psi$ such that $(H_j)_j = (G_j \circ \psi_j^{-1})_j$, because $((c_j^*)_j, \Psi)$ generates $\mathcal{C}^*$. It also holds that there exists $(\phi_j)_j \in \Phi$ such that $(G_j)_j = (G_j^* \circ \phi_j)_j$, because $((G_j^*)_j, \Phi)$ generates $\mathcal{G}^*$. Now define $L_j = (\phi_j \circ \psi_j^{-1})$ for all $j \in J$, so that $(L_j)_j \in \Phi \circ \Psi^{-1}$ holds. Furthermore,

$$(G_j^* \circ L_j)_j = (G_j^* \circ \phi_j \circ \psi_j^{-1})_j = (G_j \circ \psi_j^{-1})_j = (H_j)_j.$$

If-statement. Suppose there exists $(L_j)_j \in \Phi \circ \Psi^{-1}$ that satisfies $(G_j^* \circ L_j)_j = (H_j)_j$. Let $(\phi_j)_j \in \Phi$ and $(\psi_j)_j \in \Psi$ be such that $(L_j)_j = (\phi_j \circ \psi_j^{-1})_j$. Then $(G_j)_j := (G_j^* \circ \phi_j)_j \in \mathcal{G}^*$ because $\mathcal{G}^*$ is the maximal set generated by $((G_j^*)_j, \Phi)$, and $(c_j)_j := (c_j^* \circ \psi_j)_j \in \mathcal{C}^*$ because $\mathcal{C}^*$ is the maximal set generated by $((c_j^*)_j, \Psi)$. The data induced by $((G_j)_j, (c_j)_j)$ are, for all $t \in [\underline{t}, \bar{t}]$ and $j \in J$,

$$\begin{aligned} \hat{p}_j^0 \hat{F}_j^0(t) &= G_j((c_j^0)^{-1}(t)) \\ &= G_j^*(\phi_j((c_j^0)^{-1}(t))) \\ &= G_j^*(\phi_j(\psi_j^{-1}((c_j^{*,0})^{-1}(t)))) \\ &= G_j^*(L_j((c_j^{*,0})^{-1}(t))) \\ &= H_j((c_j^{*,0})^{-1}(t)) \\ &= p_j^0 F_j^0(t), \end{aligned}$$

where the third equality has been established in the proof of Theorem 1. The analogous argument shows that $((G_j)_j, (c_j)_j)$ induces also $p_j^1 F_j^1(t)$, hence the data is rationalizable. $\square$

For rationalizability, we need to check whether there exist functions $(L_j)_j \in \Phi \circ \Psi^{-1}$ which satisfy the implicit condition

$$G_j^*(L_j(x)) = H_j(x) \text{ for all } x \in \mathbb{R}.$$

This is easier than it may appear. For example, if both H_j and G_j^* are strictly increasing, then the candidate L_j is unique and given by $L_j = (G_j^*)^{-1} \circ H_j$, which is a bijective and strictly increasing function based on observables. It remains to be checked whether this function can be written as an admissible composition $\phi_j \circ \psi_j^{-1}$. This is always the case, for example, if Φ is the unrestricted set of all profiles of transformations, because $\Phi \circ \Psi^{-1}$ is unrestricted in that case as well. In other cases, the function L_j will be unique in some intervals and can be extended outside these intervals in a way that guarantees bijectivity and monotonicity. One then needs to check whether L_j coincides with an admissible composition $\phi_j \circ \psi_j^{-1}$ wherever it is uniquely defined.

Other cases are even easier. For example, if H_j takes the value 0 for finite values $x < 0$ but G_j^* does not, then a function L_j satisfying $G_j^* \circ L_j = H_j$ cannot be strictly increasing and hence cannot be a composition $\phi_j \circ \psi_j^{-1}$. This implies that the data is not rationalizable. Intuitively, since G_j^* extends to infinity and the data has no atoms at response time $\underline{t}$, it can only be rationalized by a chronometric function that satisfies $c_j(x) > \underline{t}$ for all $x < 0$, and hence H_j cannot reach 0.

B Robustness to Heterogeneity and Noise

As described in the main text, we model response time in the decision problem of interest as $t = c_j(x) \cdot \eta \cdot \epsilon$ and in the baseline problem as $t_b = \phi \cdot \eta$, where $\eta > 0$ captures the individual's general response speed, $\epsilon > 0$ comprises different sources of noise, and $\phi > 0$ parameterizes the baseline problem. The terms x and η may be correlated and follow a j -specific distribution. The cdf of the marginal distribution of x is denoted by $G_j(x)$ and is assumed to be continuous. The terms ϵ and ϕ may also be correlated and follow a j -specific distribution described by a joint probability measure μ_j , but they are assumed to be independent of the other variables.

In a slight deviation from the main model, we let $\underline{t} = 0$ and $\bar{t} = \infty$, as heterogeneity and noise may spread out response times to arbitrary values. The chronometric function $c_j : \mathbb{R} \rightarrow [0, \infty)$ is described by $c_j^0 : (-\infty, 0) \rightarrow [0, \infty)$ and $c_j^1 : (0, +\infty) \rightarrow [0, \infty)$, assumed to be continuous, strictly increasing/decreasing throughout, and to approach the respective limits asymptotically. Continuity of $G_j(x)$ implies that $x = 0$ is a zero probability event and allows us to leave $c_j(0)$ unspecified.

We consider normalized response times $\hat{t} = t/t_b = c_j(x) \cdot \epsilon/\phi$ and denote the cdfs of their distributions (conditional on choice) by $\hat{F}_j^i$. The model $(G_j, c_j, \mu_j)_j$ then generates the data

$$p_j^0 \hat{F}_j^0(t) = \int_{\text{supp } \mu_j} G_j((c_j^0)^{-1}(t\phi/\epsilon)) d\mu_j(\phi, \epsilon)$$

and

$$p_j^1 \hat{F}_j^1(t) = 1 - \int_{\text{supp } \mu_j} G_j((c_j^1)^{-1}(t\phi/\epsilon)) d\mu_j(\phi, \epsilon),$$

for all $t \geq 0$ and $j \in J$.

Suppose the analyst relies on the representative chronometric function $c_j^*(x) = 1/|x|$ and therefore constructs the empirical distributions

$$\hat{H}_j(x) = \begin{cases} 1 - p_j^1 \hat{F}_j^1(1/x) & \text{if } x > 0, \\ p_j^0 & \text{if } x = 0, \\ p_j^0 \hat{F}_j^0(-1/x) & \text{if } x < 0, \end{cases} \quad (19)$$

from those data, for all $j \in J$. The following result shows how these distributions, which are misspecified as they ignore heterogeneity and noise, are related to the model fundamentals.

Proposition 1. Given $(G_j, c_j, \mu_j)_j$, the distributions $(\hat{H}_j)_j$ defined in (19) can be written as

$$\hat{H}_j(x) = \int_{\text{supp } \mu_j} K_j(x\epsilon/\phi) d\mu_j(\phi, \epsilon), \quad (20)$$

where $(K_j)_j = (G_j \circ \psi_j^*)_j$ with

$$\psi_j^*(x) = \begin{cases} (c_j^1)^{-1}(1/x) & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ (c_j^0)^{-1}(-1/x) & \text{if } x < 0, \end{cases}$$

for all $j \in J$.

Proof. Consider any $j \in J$. For $x = 0$ we have

$$\int_{\text{supp } \mu_j} K_j(x\epsilon/\phi) d\mu_j(\phi, \epsilon) = K_j(0) = G_j(0) = p_j^0 = \hat{H}_j(0).$$

For any $x < 0$ we have

$$\begin{aligned} \int_{\text{supp } \mu_j} K_j(x\epsilon/\phi) d\mu_j(\phi, \epsilon) &= \int_{\text{supp } \mu_j} G_j((c_j^0)^{-1}(-(1/x)(\phi/\epsilon))) d\mu_j(\phi, \epsilon) \\ &= p_j^0 \hat{F}_j^0(-1/x) = \hat{H}_j(x). \end{aligned}$$

For any $x > 0$ we analogously have

$$\begin{aligned} \int_{\text{supp } \mu_j} K_j(x\epsilon/\phi) d\mu_j(\phi, \epsilon) &= \int_{\text{supp } \mu_j} G_j((c_j^1)^{-1}((1/x)(\phi/\epsilon))) d\mu_j(\phi, \epsilon) \\ &= 1 - p_j^1 \hat{F}_j^1(1/x) = \hat{H}_j(x), \end{aligned}$$

which establishes (20). □

Proposition 1 extends the main insight from the proof of Theorem 1 to the case with noise. Indeed, if there is no noise (i.e., ϵ/ϕ is fixed at 1), then it implies that $\hat{H}_j = K_j = G_j \circ \psi_j^*$, where ψ_j^* corresponds to ψ_j^{-1} in the proof of Theorem 1.

In the general case, the result can be applied to check whether $(\hat{H}_j)_j$ is suitable for testing the hypothesis that $(G_j)_j$ satisfies a property **P**. The first requirement is that **P** of $(G_j)_j$ is invariant to ψ_j^* , so that $(K_j)_j$ also satisfies **P**. Without knowledge of the chronometric function beyond monotonicity, this requires invariance to all increasing transformations. With the knowledge of symmetry, it requires invariance to symmetric transformations, and

so on. The second (and crucial) requirement is invariance of $\mathbf{P}$ of $(K_j)_j$ to the multiplicative convolutions (20), so that $(\hat{H}_j)_j$ inherits $\mathbf{P}$.

To illustrate the argument, consider the property that a single distribution G is strictly increasing (full support). This property is invariant to all strictly increasing transformations, and furthermore it is invariant to (20) for arbitrary noise measures μ , based on the observation that increasing x increases $K_j(x\epsilon/\phi)$ pointwise for each (ϕ, ϵ) in the integral. Hence strict monotonicity of G carries over to strict monotonicity of $\hat{H}$ without any additional restrictions on the model. Consequently, if $\hat{H}$ is not strictly increasing, the hypothesis that G is strictly increasing can be rejected.

As a second illustration, consider the property of a pair (G_1, G_2) of distributions that $G_1(-x) + G_1(x) \leq G_2(-x) + G_2(x)$ for all $x \geq 0$, which implies that the mean of G_1 is larger than the mean of G_2 and which we use in our applications in Section 3. This property is invariant to transformations that are symmetric around zero and identical for the two groups. The property is then also invariant to (20) for noise measures μ_j that are identical for the two groups, again based on a pointwise argument in the integrals. Consequently, $(\hat{H}_1, \hat{H}_2)$ can be used to test the property under the assumption of symmetric and common chronometric functions and common noise distributions.

C Ranking of Means

In this appendix, we show that the inequality $G_A(x) + G_A(-x) \leq G_B(x) + G_B(-x)$ for all $x \in \mathbb{R}_+$ implies that the mean of G_A , denoted μ_A , is larger than the mean of G_B , denoted μ_B (assuming that both means exist). Using the fact that

$$\mu_j = - \int_{-\infty}^0 G_j(x) dx + \int_0^{+\infty} [1 - G_j(x)] dx,$$

we have

$$\begin{aligned} \mu_A - \mu_B &= \int_{-\infty}^0 [G_B(x) - G_A(x)] dx + \int_0^{+\infty} [1 - G_A(x) - 1 + G_B(x)] dx \\ &= \int_0^{+\infty} [G_B(-x) - G_A(-x)] dx + \int_0^{+\infty} [G_B(x) - G_A(x)] dx \\ &= \int_0^{+\infty} [G_B(x) + G_B(-x) - G_A(x) - G_A(-x)] dx \geq 0. \end{aligned}$$

D Further Applications

This appendix develops further applications of Theorem 1 with distributional properties that arise in a variety of empirical settings. For each property, we identify the relevant invariance class and characterize the conditions under which detection is feasible using response times. Although we do not implement these applications empirically, the results provide a template for future work.

D.1 Inequality

Our method can be used to study inequality or dispersion of a latent variable, which has applications across multiple fields. For instance, within the literature on subjective well-being, there is a substantial interest in understanding the inequality of happiness (e.g., Stevenson and Wolfers, 2008). Similarly, researchers have studied societal polarization by measuring the dispersion of individual attitudes towards social and political issues (DiMaggio, Evans and Bryson, 1996; Evans, 2003). In the context of market competition, the spread of consumer preferences has direct implications for the optimal pricing and advertising strategies of firms (Johnson and Myatt, 2006; Hefti, Liu and Schmutzler, 2022). Unfortunately, standard measures of inequality like the Gini index require cardinal information, which makes their application to ordered response data questionable (see the discussion in Dutta and Foster, 2013). Response times may serve as a source of cardinal information even when decisions are binary, such as whether to buy or not buy a product (Cotet, Zhao and Krajbich, 2025).

The Lorenz curve is a convenient graphical representation of a distribution's inequality, and interesting measures of inequality like the Gini index are based on the Lorenz curve (Atkinson, 1970; Cowell, 2011). For any distribution G , the Lorenz curve is defined by

$$L(q, G) = \frac{\int_0^q G^{-1}(x)dx}{\int_0^1 G^{-1}(x)dx} \text{ for all } q \in [0, 1],$$

where $G^{-1}(x) := \inf\{y \mid G(y) \geq x\}$ denotes the left inverse of G . In the context of subjective well-being, $L(q, G)$ could be understood as the proportion of total happiness allocated to the least happy 100 q percent of the population. How far the Lorenz curve falls below the 45-degree line is an indication of how unequal the distribution is. The curve is invariant to all linear transformations of the distribution G . Therefore, if we plot the Lorenz curve for an empirical function H constructed as in (5) with any representative function c^* , exactly this Lorenz curve (and any measure based on it like the Gini index) is detected for the class of chronometric functions that are linearly generated from c^* . One can repeat this procedure

for various different representative functions c^{k*} to obtain bounds on the true Lorenz curve for larger sets of chronometric functions.²⁶

Beyond measuring inequality within a single population, some applications require comparing inequality across groups. The subjective well-being literature has been interested in how inequality of happiness changed over time and across nations (Kalmijn and Veenhoven, 2005; Stevenson and Wolfers, 2008; Dutta and Foster, 2013). There is also a great interest in political polarization trends (DiMaggio, Evans and Bryson, 1996; Evans, 2003). Furthermore, firms have an interest in learning about changes in the spread of consumer preferences, to optimally adapt their pricing and advertising strategies (Johnson and Myatt, 2006; Hefti, Liu and Schmutzler, 2022). Our method applies to such comparative analyses as well.

Consider first the problem of detecting whether G_1 has a smaller variance than G_2 , which is a convenient way of ranking the dispersion of distributions due to its ability to provide a complete order. Just as for the ranking of means, this property is invariant only to transformations (ψ_1, ψ_2) that are linear and additionally satisfy $\psi_1 = \psi_2$. Comparing instead the inequality based on Lorenz curves has the benefit of avoiding direct scale comparisons, as the two distributions' Lorenz curves are invariant to transformations (ψ_1, ψ_2) that are linear but not necessarily identical. For example, we can say that G_1 Lorenz-dominates G_2 if $L(q, G_1) \geq L(q, G_2)$ holds for all $q \in [0, 1]$ (see Shaked and Shanthikumar, 2007). Then, under the assumption of knowing the chronometric functions up to group-specific linear transformations, detecting a Lorenz-dominance relationship between G_1 and G_2 (or the absence thereof) is equivalent to checking whether the property holds for the respective empirical functions (H_1, H_2) .

Another measure that can be used for assessing the relative dispersion of distributions is the concept of single-crossing dominance (Diamond and Stiglitz, 1974; Hammond, 1974). Formally, G_1 single-crossing dominates G_2 if there exists x^* such that $G_1(x) \leq G_2(x)$ if $x \leq x^*$ and $G_1(x) \geq G_2(x)$ if $x \geq x^*$. Intuitively, the condition reflects that G_1 assigns less probability weight to values at the tails of the distribution compared to G_2 . Although the property of single-crossing dominance is sensitive to group-specific transformations, it has the advantage of being robust to non-linear transformations. Just like FOSD, the property (and its violation) is invariant to all profiles of transformations (ψ_1, ψ_2) that satisfy $\psi_1 = \psi_2$. Therefore, according to Theorem 1, single-crossing dominance is detected for a relatively large class of chronometric functions if H_1 single-crossing dominates H_2 , where (H_1, H_2) are constructed from data using (5), and a violation thereof is detected otherwise. The condition can again be expressed directly in terms of the data. It requires that $p_1^i F_1^i(t)$ and $p_2^i F_2^i(t)$

²⁶Analogous arguments apply to other distributional properties that are invariant to linear transformations, like skewness or kurtosis.

cross at most once for one of the two responses $i \in \{0, 1\}$ and not at all for the other response.

Finally, it is known that either Lorenz-dominance or single-crossing dominance imply second-order stochastic dominance (SOSD) when G_1 has a higher mean than G_2 (Thistle, 1989). Therefore, combining several of the results derived so far allows us to detect SOSD.

D.2 Likelihood-Ratio Dominance

Consider the property that G_1 likelihood-ratio dominates G_2 , defined by the inequality

$$(G_1(x) - G_1(x''))(G_2(x) - G_2(x')) \leq (G_1(x) - G_1(x'))(G_2(x) - G_2(x''))$$

for all $x'' < x' < x$ (see Wang and Lehrer, 2024). If G_1 and G_2 are absolutely continuous, the property can be equivalently expressed in terms of their density functions g_1 and g_2 as $g_1(x')g_2(x) \leq g_1(x)g_2(x')$ for any $x' < x$ (see Shaked and Shanthikumar, 2007). Likelihood-ratio dominance, which is stronger than FOSD, has proven useful in a range of economic applications that involve monotone comparative statics under uncertainty (Milgrom, 1981; Athey, 2002). Suppose that the goal is to maximize the expected value of a function $\pi(y, x)$ by choice of y when x is distributed according to G_j . If $\pi(y, x)$ satisfies a single-crossing property and G_1 likelihood-ratio dominates G_2 , then the optimal choice of y is larger for G_1 than for G_2 (under appropriate technical conditions, see Athey, 2002). For example, a firm may introduce a new product in two different countries, each of which is characterized by a distribution of consumer types. Higher types imply a higher demand for the product in a way that the firm's profit $\pi(p, x)$ as a function of price p and type x is single-crossing. If the firm's market research (e.g., a survey that asks potential consumers whether they like the product) allows the detection of likelihood-ratio dominance of the type distributions of the two countries, the firm knows in which country to charge a higher price. Other applications of the concept include optimal investment decisions and bidding in auctions.

Like FOSD, likelihood-ratio dominance (and its violation) is invariant to all profiles of transformations (ψ_1, ψ_2) that satisfy $\psi_1 = \psi_2$. We thus proceed as for FOSD by constructing (H_1, H_2) from data and then verifying whether H_1 likelihood-ratio dominates H_2 . For example, checking the respective inequalities for all $x'' < x' < x \leq 0$ can be expressed in terms of the data as

$$(p_1^0 F_1^0(t) - p_1^0 F_1^0(t''))(p_2^0 F_2^0(t) - p_2^0 F_2^0(t')) \leq (p_1^0 F_1^0(t) - p_1^0 F_1^0(t'))(p_2^0 F_2^0(t) - p_2^0 F_2^0(t''))$$

for all $t'' < t' < t$. The full condition is particularly easy to express when $p_j^i > 0$ and the response time distributions F_j^i admit strictly positive densities f_j^i . In that case, we detect

likelihood-ratio dominance if the empirical likelihood-ratio

$$\frac{p_1^i f_1^i(t)}{p_2^i f_2^i(t)}$$

is weakly increasing in t for $i = 0$ and weakly decreasing in t for $i = 1$, and

$$\frac{p_1^0 f_1^0(\bar{t})}{p_2^0 f_2^0(\bar{t})} \leq \frac{p_1^1 f_1^1(\bar{t})}{p_2^1 f_2^1(\bar{t})}$$

holds. Otherwise, a violation of likelihood-ratio dominance is detected.

Our approach extends analogously to the detection of hazard-rate dominance and reversed hazard-rate dominance, both of which lie between likelihood-ratio dominance and FOSD, and which are also commonly used in applied settings (e.g., Kiefer, 1988; Maskin and Riley, 2000; Wang and Lehrer, 2024).

D.3 Unimodality

Consider the property that the distribution is unimodal with mode at zero, i.e., G is convex below zero and concave above zero, and strictly so except when $G(x) \in \{0, 1\}$. Unimodality is of interest once more in political applications where G describes the distribution of political attitudes in a population. Unimodality reflects a centered population where more extreme positions receive less support. An ongoing debate in political science concerns whether political attitudes follow unimodal distributions or are polarized and better described by bimodal distributions (see Lelkes, 2016; Vaeth, 2023). Analysts often want to learn about these properties from survey responses, but as Vaeth (2023) points out, ordinal responses are inadequate to test for properties like uni- or bimodality of the underlying distribution.

Unimodality is invariant to (sigmoid) transformations that satisfy $\psi(0) = 0$ and are weakly convex below zero and weakly concave above zero. Starting from a representative chronometric function c^* , such transformations allow us to generate all chronometric functions which are weakly “more convex” than c^* on $\mathbb{R}_-$ and on $\mathbb{R}_+$ (because c^* is increasing on $\mathbb{R}_-$ but decreasing on $\mathbb{R}_+$). We will use this insight in a slightly different way than before. Assume that the observed cdfs F^i are strictly increasing on $[\underline{t}, \bar{t}]$. Then, construct from the

data the representative chronometric function

$$c^*(x) = \begin{cases} \underline{t} & \text{if } 1 < x, \\ (F^1)^{-1}(1 - x) & \text{if } 0 < x \leq 1, \\ (F^0)^{-1}(1 + x) & \text{if } -1 \leq x \leq 0, \\ \underline{t} & \text{if } x < -1, \end{cases} \quad (21)$$

which reaches $c(x) = \underline{t}$ for finite absolute values of x . With this function, the empirical distribution H becomes

$$H(x) = \begin{cases} 1 & \text{if } 1 < x, \\ 1 - p^1 + p^1 x & \text{if } 0 < x \leq 1, \\ p^0 + p^0 x & \text{if } -1 \leq x \leq 0, \\ 0 & \text{if } x < -1, \end{cases}$$

which is piecewise linear. Applying any strictly sigmoid transformation to (21) generates a chronometric function that also attains $c(x) = \underline{t}$ for finite absolute values of x but is piecewise more convex, resulting in a unimodal H . By the same logic, applying any inverse-sigmoid transformation generates a piecewise more concave chronometric function and a distribution H that is not unimodal. The constructed function (21) therefore delimits sets of chronometric functions for which we can detect or reject unimodality. It follows from Theorem 1 that unimodality is detected for all those functions that are more convex than (21) and rejected for those that are more concave. This approach does not cover all possible chronometric functions but may yield expressive results. For example, if (21) plotted from the data is already strongly convex, then it appears unlikely that the true chronometric function is even more convex, and we may be able to reject the assumption of unimodality of the distribution.

D.4 Correlation

Our last application expands upon the baseline model by investigating the correlation between a latent variable $x \in \mathbb{R}$ (e.g., happiness, political attitude, or willingness to pay) and an observable variable represented by the index $j \in \mathbb{R}$ (e.g., income, social media usage, or experimental treatments). Such an application can be of value in the context of opinion surveys on topics that are sensitive and where subjects hesitate to provide certain answers that conflict with social norms (e.g., Coffman, Coffman and Ericson, 2017), which can result

in limited variation in the response data. One may compensate for the lack of power of traditional analysis by replacing response variation with variation in response times. This can also be useful for experimental economists who are interested in the effect of different treatments on subjects' behavior. A null result in behavior does not necessarily establish the absence of an effect. If choices are discrete, such a null result can arise when the effect exists but is not large enough to shift behavior, given the parameters chosen by the experimenter. The effect may still be detectable in response times. The use of response time data for these purposes has already been proposed by Konovalov and Krajbich (2019).

We define a cumulative distribution function Γ over the indices $j \in J \subseteq \mathbb{R}$, which is observable, and interpret $(G_j)_j$ as the conditional distributions of x given each j . The joint distribution of (x, j) is fully determined by $(G_j)_j$ and Γ . Further, just like in the baseline model, given observed data $(p_j, F_j)_j$ and a representative profile $(c_j^*)_j$ of chronometric functions, we can derive empirical distribution functions $(H_j)_j$ as defined in (5). These functions together with Γ also give rise to a well-defined joint distribution of (x, j) .

A first attempt to quantify the association between x and j is to employ the standard Pearson correlation coefficient

$$\rho = \frac{Cov(x, j)}{\sqrt{Var(x) Var(j)}}, \quad (22)$$

where Cov indicates the covariance function and Var indicates the variance function. The value of this coefficient is invariant to all profiles $(\psi_j)_j$ of positive affine transformations that are identical for all $j \in J$. Consequently, when the chronometric functions are known up to identical linear transformations, we can ascertain the exact correlational pattern by computing (22) for the joint distribution of (x, j) constructed via the functions $(H_j)_j$ and the marginal distribution Γ .

To circumvent the linearity restriction, one might opt for measuring the rank correlation between x and j , which is particularly natural when x and j are ordinal variables (Kendall, 1955). Intuitively, the rank of a variable is preserved under any monotone transformation, thereby allowing for more robust detection. For example, consider Spearman's (1904) rho, defined as

$$\rho_s = \frac{Cov(G(x), \Gamma(j))}{\sqrt{Var(G(x)) Var(\Gamma(j))}},$$

where the function G represents the marginal distribution of x and is given by

$$G(x) = \int_J G_j(x) d\Gamma(j) \text{ for all } x \in \mathbb{R}.$$

Since the rank of x within the population remains unchanged under any profile $(\psi_j)_j$ of strictly increasing transformations that are identical across j , so does the value of Spearman's rho. A similar observation holds for another rank-based correlation measure, Kendall's (1955) tau, which is defined as

$$\rho_\tau = \mathbb{E} \left[\mathbb{1}_{\{(x-x')(j-j') > 0\}} \right] - \mathbb{E} \left[\mathbb{1}_{\{(x-x')(j-j') < 0\}} \right],$$

where (x', j') is distributed independently of (x, j) but with the same joint distribution. Consequently, the rank correlation patterns for our variables of interest are detected under a fairly general class of chronometric functions, whenever they hold for the joint distribution induced by the appropriate functions $(H_j)_j$ and Γ .²⁷

²⁷The invariance property shared by Spearman's rho and Kendall's tau is related to the fact that they both depend only on the bivariate Copula of the two random variables, which is invariant to monotone transformations; see, e.g., Fan and Patton (2014) and Haugh (2016). The population versions of Spearman's rho and Kendall's tau that we adopt here were taken from Haugh (2016).