Economics Working Paper Series

2024/006

Identifying Nontransitive Preferences

Carlos Alós-Ferrer, Ernst Fehr and Michele Garagnani

The Department of Economics
Lancaster University Management School
Lancaster LA1 4YX
UK

Identifying Nontransitive Preferences*

Carlos Alós-Ferrer†[1], Ernst Fehr[2], and Michele Garagnani[3]

[1]Lancaster University Management School
[2]Department of Economics, University of Zurich
[3]Department of Finance, University of Melbourne

This Version: May 13, 2024

Abstract

Transitivity is perhaps the most fundamental axiom in economic models of choice. The empirical literature has regularly documented violations of transitivity, but these violations pose little problem if they are simply a result of somewhat-noisy decision making and not a reflection of the deterministic part of individuals' preferences. However, what if transitivity violations reflect genuinely nontransitive preferences? And how can we separate nontransitive preferences from noise-generated transitivity violations–a problem that so far appears unresolved? To tackle these fundamental questions, we develop a theoretical framework which allows for nontransitive choices and behavioral noise. We then derive a non-parametric method which uses response times and choice frequencies to distinguish genuine (and potentially nontransitive) preferences from noise. We apply this method to two different datasets, demonstrating that a substantial proportion of transitivity violations reflect genuinely nontransitive preferences. These violations cannot be accounted for by any model using transitive preferences and noisy choices.

JEL Classification: D01 · D81 · D87 · D91

Keywords: Transitivity · Stochastic choice · Preference Revelation

*We thank Stefano Dellavigna, Georg Nöldeke, Rob Ranyard, Michael Regenwetter, Ulrich Schmidt, and Eyal Winter for helpful comments. We also thank seminar participants at the Universities of Basel, Bonn, Ca' Foscari of Venice, Lancaster, Lausanne, Maastricht, Melbourne, Monash, New York University at Abu Dhabi, Southampton, and Sydney, and conference audiences at the Foundations of Utility and Risk (Ghent), Economic Science Association (Bologna), and Australasian Mathematical Psychology Conference (Perth).

†Corresponding author. Lancaster University Management School, LA1 4YX Lancaster (U.K.). E.mail: c.alosferrer@lancaster.ac.uk

1 Introduction

The economic approach to decisions builds upon the assumption that choices can be represented by (complete) transitive binary relations, that is, preferences. Transitivity is hence, arguably, the most fundamental assumption behind economic models of choice. Yet, the empirical literature has regularly documented systematic violations of transitivity in the form of cyclical choices where, for example, a is chosen over b, b is chosen over c, and c is chosen over a (e.g., Tversky, 1969; Loomes et al., 1989, 1991; Starmer, 1999, 2000; Humphrey, 2001).

The interpretation of this empirical evidence is strongly contested. The main argument is that choice is stochastic, and hence it is possible to observe nontransitive choices even though preferences are transitive, because actual choices are noisy (Iverson and Falmagne, 1985; Sopher and Gigliotti, 1993; Birnbaum, 2020). As Birnbaum and Schmidt (2010) observed, *"[a] problem that has frustrated previous research has been the issue of deciding whether an observed pattern represents 'true violations' of transitivity or might be due instead to 'random errors.'"* In other words, observed transitivity violations could be explained by, for example, random utility models which postulate a transitive binary relation plus a noise term (McFadden, 1974, 2001; Anderson et al., 1992). The current literature has long been at an *impasse* due to the impossibility of disentangling preferences from noise.

In this contribution, we show how to disentangle preferences and noise to examine whether cyclical choices are due to noise or true evidence of genuinely nontransitive preferences. In this sense, we follow the views of Machina (1985), who already argued that stochastic nontransitivity is compatible with transitive choice plus noise (or randomization), and stated that *"the 'proper' notion of intransitivity [...] ought to be the phenomenon of intransitive underlying preferences over the [options] themselves."* Specifically, our theoretical approach and empirical analyses allow us to show the existence of nontransitive choice patterns which cannot be explained by *any* model built upon transitive preferences and noisy choices. To do so, we consider not a specific model, but rather a general framework encompassing any model (transitive or not) of noisy choice. This is important because, were we to consider a specific model, it would be in principle possible to argue that a different model might explain the transitivity violations we find while assuming transitivity. In contrast, our approach settles the debate and provides conclusive evidence for the existence of nontransitive preferences.

To accomplish this objective, we extend recent results by Alós-Ferrer et al. (2021), which use response times to reveal preferences even when choices alone cannot do so, but do not require any particular model or parametric assumption. Our approach thus allows for preference *revelation* (as opposed to estimation, which requires a specific model) even when the underlying binary relation is nontransitive. We then apply the results to two existing datasets by Davis-Stober et al. (2015) and Kalenscher et al. (2010), hereafter DSBC and KTHDP, respectively. Applying the new results, we find that both datasets exhibit substantial percentages of transitivity violations in the underlying preferences, independently of any model of noise.

The revelation result we use is based on two robust empirical regularities of choices and response times arising from psychology and neuroscience. The first is that easier choice problems are more likely to elicit correct responses than harder problems. This *psychometric effect* is one of the most robust facts in all of psychology (Cattell, 1893; Dashiell, 1937; Laming, 1985; Klein, 2001; Wichmann and Hill, 2001). It has also been established for economic decisions, with evidence dating back to Mosteller and Nogee (1951) and including Fudenberg et al. (2018) and Alós-Ferrer and Garagnani (2022a,b). The second regularity is that easier choice problems take less time to respond to than harder problems. There is overwhelming evidence for this *chronometric effect* in a wide variety of domains, starting with Cattell (1902), Moyer and Landauer (1967), Moyer and Bayer (1976), and Dehaene et al. (1990). The finding extends to preferential choices (e.g., Dashiell, 1937), and a growing number of contributions have demonstrated it for intertemporal choice (Chabris et al., 2009), social preferences (Krajbich et al., 2015), and decisions under risk (Moffatt, 2005; Konovalov and Krajbich, 2019; Alós-Ferrer and Garagnani, 2022a,b).[1]

A simplified intuition for our analysis and our results is as follows. By the effects described above, slow decisions are often more noisy than fast ones, because they correspond to choices where the decision maker is closer to indifference (informally speaking) and noise is more likely to offset the actual preference. Hence, if received transitivity violations were mostly due to noise, they should be associated with at least some slow choices along the corresponding cyclical patterns. By the same logic, chains of fast decisions should almost never lead to transitivity violations. In contrast, if received transitivity violations mostly reflect genuinely nontransitive preferences, we should find substantial numbers of violations where all involved choices are fast.

A simple reanalysis of the datasets of DSBC and KTHDP illustrates this rough intuition. For this exercise, a decision for a given participant is "slow" ("fast") if its average response time across all repetitions of the decision is above (below) the participant's median response time across all decisions. We then rely on the standard measure of transitivity violations used in the literature, violations of *Weak Stochastic Transitivity* (WST; see below for details). In DSBC, on average, participants violated WST along 19.53 potential cycles out of 95 (20.56%). Of those, on average 10.62 were such that all decisions along the cycle were fast in the sense defined above, and the remaining 8.91 were such that at least one of them was slow. That is, on average across participants, 54.38% of all transitivity violations are, intuitively, hard to explain as the result of noise. This represents 11.18% of all potential cycles in the experiment. For KTHDP, the situation is even more extreme. Participants on average violated WST along 47.19 potential

[1]These effects were originally documented in discrimination tasks, where a decision is hard when the difference between two stimuli is small. As argued by Fudenberg et al. (2018), large error rates and long response times in this case might reflect the difficulty in separating the values of the options.

cycles out of 306 (15.42%), and for 42.59 of those, all decisions along the cycle were fast (and only 4.59 involved one or more slow decisions). Hence, in this experiment 90.25% of all transitivity violations are hard to explain as the result of noise, representing 13.92% of all the potential cycles in the experiment.

In our analysis, we go beyond this simple intuition. The theoretical approach that we develop allows us to identify *Revealed Transitivity Violations* (RTVs) in datasets including repeated choices and response times. RTVs are cyclical patterns of choices such that, for each choice pair along the cycle and for *any* model of preference-based choice (transitive or not) including noise (no matter which assumptions on the latter are imposed, e.g. symmetric or not), the data reveals that the underlying preference is as specified in the cycle. Hence, the observed preference cycle can only be explained by a genuinely nontransitive preference, and not by choice noise.

This approach strongly differs from the previous literature, which has concentrated on violations of WST and related properties (Tversky, 1969). Denoting by $p(x, y)$ the proportion of x choices from the pair $\{x, y\}$, a WST violation is a pattern in the data where $p(a, b) \geq 1/2$ and $p(b, c) \geq 1/2$, but $p(a, c) < 1/2$. In contrast, an RTV is a pattern in the choice and response time data which reveals preferences such that $a \succeq b$, $b \succeq c$, but $c \succ a$. We will show that RTVs are more demanding than WST violations in the sense that every RTV implies a violation of WST (see Section 4.1), but the converse is not true. This is because violations of WST could be explained through models of noisy choice where preferences are transitive but noise is asymmetrically distributed, while RTVs cannot be explained by *any* model of noisy choice.

Our objective is to falsify the transitivity hypothesis in arbitrary models with noisy choices and hence be able to demonstrate the existence of genuinely nontransitive preferences without relying on any particular model specification. We build upon the (deterministic) framework developed in the seminal paper of Shafer (1974). In a standard utility model, x is (weakly) preferred to y if and only if $u(x) - u(y) \geq 0$, where u is a utility function. Shafer (1974) replaces utilities with general two-variable functions $v(x, y)$, which can be thought of as "strength of preference" (e.g. Dyer and Sarin, 1982; Fishburn, 1988), such that x is (weakly) preferred to y if and only if $v(x, y) \geq 0$. This allows for nontransitive choices, as $v(x, y) > 0$ and $v(y, z) > 0$ do not necessarily imply that $v(x, z) > 0$. The approach encompasses models allowing for nontransitive choices such as (generalized) regret theory (Loomes and Sugden, 1982, 1987; Bleichrodt and Wakker, 2015), salience theory (Bordalo et al., 2012), and Skew-Symmetric-Bilinear utility (SSB; Fishburn, 1984a,b,c), which postulate specific functions $v(x, y)$ capturing particular phenomena (e.g., regret or salience).

We then enrich this framework allowing for noise along the lines of a generalized version of random utility models (RUMs). In a RUM, x is chosen over y if and only if $u(x) - u(y) + \varepsilon_{xy} > 0$, where ε_{xy} is a pair-specific noise term. This encompasses standard additive RUMs, but also random parameter models as in, e.g., Loomes and Sugden (1998) or Apesteguía and Ballester (2018). In our *Random Choice Models* (RCMs), x is chosen over y if and only if $v(x, y) + \varepsilon_{xy} > 0$. Hence, RCMs encompass noisy-but-transitive models as RUMs while allowing for genuinely nontransitive preferences. RCMs also encompass nontransitive-but-deterministic models as regret theory or salience theory while allowing for noisy choice.

We work in the universe of RCMs and derive a (possibly-nontransitive) preference revelation result extending the main result of Alós-Ferrer et al. (2021), which we then apply to the data. Our approach provides conditions which, if fulfilled, reveal the underlying preference within a pair independently of any assumptions on the noise. Those revealed preferences can in turn reveal nontransitive cycles. That is, contrary to WST and related approaches, we do not look for violations of certain implied conditions (on choice frequencies only), but rather examine when genuinely nontransitive preferences are revealed by the choice and response time data. In this sense, an RTV does not just suggest that the data might violate transitivity: it actually reveals nontransitive preferences behind the data. Thus, we provide a framework where noise can be disentangled from underlying (potentially nontransitive) preferences. If a nontransitive preference cycle is revealed, we can conclude that there exists no model of noisy choice relying on transitive preferences which can explain the data. Hence, our results allow to demonstrate the existence of genuinely nontransitive preferences.

The previous literature is characterized by a back-and-forth between contributions showing empirical violations of WST and related criteria, and responses arguing that those might be explained by models taking noise into account (see Online Appendix A for a more detailed review of this literature). Tversky (1969) reported WST violations, but Iverson and Falmagne (1985) reanalyzed the data and argued that evidence was compatible with transitive preferences plus noise. Loomes et al. (1989, 1991) invoked transitivity violations as a potential explanation of anomalies in risky choice, but Sopher and Gigliotti (1993) replicated their experiments and found choices to be captured by a structural model with transitive preferences and random errors.

Regenwetter et al. (2010, 2011) argued that violations of WST might be caused by *stochastic preferences* (Block and Marschak, 1960), that is, probability distributions over transitive preferences. They suggested to analyze possible violations of transitivity through violations of the Triangle Inequality instead: $p(x, y) + p(y, z) - p(x, z) \leq 1$. This property must be satisfied by any stochastic preference. Using both WST and the Triangle Inequality, Butler and Pogrebna (2018) found evidence for nontransitive preferences, but Birnbaum (2020, 2023) argued that those choice patterns could be explained by allowing both for stochastic preferences and additional choice errors. Many other contributions in the literature have exhibited choice patterns possibly reflecting nontransitive preferences in multiple domains (see Online Appendix A). We refer the reader to Starmer (2000) and to the recent review by Ranyard et al. (2020).

Our contribution, however, is a major conceptual departure from this previous literature. Our approach does not rest on any specific model, or on conducting a "horse race" to see which of a given set of models explains data better (a common approach used, e.g., by Ranyard et al., 2020). On the contrary, we identify transitivity violations which cannot be explained by *any* noisy-choice model which assumes transitive preferences, or distributions over them. The class of models discarded by our analysis is far more general than any class previously considered in the literature. In particular, it does not assume that preferences are stable, since it encompasses RUMs which are equivalent to stochastic preferences, i.e. probability distributions over different preferences. This is important because, as mentioned above, it has been argued that violations of transitivity might just be due to decisions being best described by a distribution over transitive preferences (Regenwetter et al., 2011). Further, RCMs do *not* assume that noise is an additional term added to utilities, as in standard RUMs. In particular, it includes models which *cannot* be represented in that fashion, because noise terms are pair-specific and violate the Axiom of Revealed Stochastic Preference (McFadden, 1974; McFadden and Richter, 1990). That is, if our method reveals a transitivity violation, the interpretation is not that a certain nontransitive model "fits the data better." Rather, the interpretation is that there exists *no* model derived from transitive preferences (stable or not, deterministic or stochastic) and noise (additive or not, alternative-specific or pair-specific) which can explain the data.

Our main empirical point is the demonstration that a substantial amount of genuine transitivity violations can be found in standard economic choice tasks, hence rejecting the hypothesis that choices can be represented by transitive preferences plus behavioral noise. This requires datasets where subjects made the same choice multiple times (as in any experiment focusing on WST violations) *and* where response times were explicitly and reliably measured, as in the datasets of DSBC and KTHDP. It is important to note that none of these datasets was collected with our approach in mind, and hence they also serve as a demonstration of the applicability of our techniques.

In addition to demonstrating revealed transitivity violations in the data, we also compare them to violations of WST in both datasets. Not all violations of WST are genuine violations of transitivity, and hence our approach provides a better estimate of the extent of nontransitive preferences. In the Online Appendix E we also examine which choice patterns give rise to nontransitivities more often in both datasets, and hence which might be the mechanisms underlying genuinely nontransitive preferences. The most frequent nontransitivities appear to involve chains of decisions accepting small changes in the characteristics of the options which are then undone when considered together. People seem to accept series of small tradeoffs in a way that does not scale up. For example, they repeatedly accept small decreases in monetary payoffs in exchange for small increases in the probability of a payoff, until a point is reached where they accept a large decrease in probability in exchange for a large increase in the monetary payoff, bringing them back to the starting point. Crucially, however, this is unrelated to the traditional idea of "nontransitive indifference." This argument points out that apparent nontransitivities might be observed when decision makers are close to indifference, because (by the psychometric effect) noise is maximal in this case. However, our approach disentangles preferences from noise, hence this kind of apparent nontransitivities will by definition not result in revealed transitivity violations. In other words, "nontransitive indifference" produces apparent nontransitivities due to noise, while the typical nontransitivities we observe (chains of small changes which do not scale up) are genuine and not due to noise derived from indifference.

The paper is structured as follows. Section 2 reviews the deterministic models which allow for transitivity violations (Section 2.1), explains why the models we consider are more general than standard (additive) random utility models (Section 2.2), and presents our (nontransitive) preference revelation result based on response times (Section 2.3). Section 3 presents our empirical analysis of two existing lottery-choice datasets and applies the techniques to uncover the extent of revealed transitivity violations. Section 4 compares those to violations of WST. Section 5 concludes. Additionally, the Online Appendix collects robustness checks, examines the most frequent nontransitive choice patterns in the data, and also presents a more detailed discussion of the previous empirical literature on transitivity violations and deterministic models allowing for nontransitivities.

2 Distinguishing Noise from Nontransitive Preferences

To test whether choices are transitive, one needs to allow for the possibility that they are not. Following Shafer (1974) and others, we refer to a complete but not necessarily transitive binary relation as a *nontransitive preference*. We build up the framework in three steps. First (Subsection 2.1), we review deterministic models of nontransitive choice, encompassing skew-symmetric bilinear (SSB) utility theory, generalized regret theory, and salience theory. Second (Subsection 2.2), we review how to incorporate noise into models of choice. We provide a generalization of standard, additive, random utility models, and show that the standard conditions used in the literature to test for violations of transitivity are insufficient. Third (Subsection 2.3), we bring both (potentially nontransitive) preferences and general models of noise together and proceed to extend the (already generalized) random utility models to allow both for nontransitivities which are simply due to noise and those which are due to underlying nontransitive preferences.

2.1 Deterministic Models of Nontransitive Preferences

If transitivity does not hold, choices can not be represented by utility functions. It is, however, possible to represent nontransitive binary relations on a set X through real-valued, two-argument functions as follows. Consider a skew-symmetric function $v : X^2 \mapsto \mathbb{R}$, i.e. $v(x, y) = -v(y, x)$ for all $x, y \in X$. We say that a nontransitive preference $\succeq$ on X is represented by a function $v : X^2 \mapsto \mathbb{R}$ if, for all $x, y, \in X$, $v(x, y) \geq 0$ holds if and only if $x \succeq y$. For Euclidean spaces, Shafer (1974) proved that every strictly convex and continuous nontransitive preference can be represented by a continuous, skew-symmetric function as above. This is a natural generalization of representation results for transitive preferences, in which case one can set $v(x,y) = u(x) - u(y)$ for a utility function u. Interestingly, the function v has been interpreted as a "strength of preference" (see, e.g. Fishburn, 1988, Chapter 3.9 and ff.), with values of $v(x,y)$ close to zero indicating a difficult decision (the decision maker is close to indifference).

A large number of important models are based on this idea (see Online Appendix B for further details). A classical example is the *additive difference model* (Tversky, 1969), which postulates a specific function of the form $v(x,y)$ for multidimensional alternatives, with the explicit purpose of studying nontransitivities. For the case of lottery choice, Fishburn (1982, 1984b, 1986) studied skew-symmetric bilinear (SSB) representations (originally proposed by Kreweras, 1961), which are based on a function of the form $v(x,y)$ with the added requirement of linearity in both arguments, yielding a natural generalization of expected utility theory.

A number of prominent models have incorporated general behavioral phenomena which result in nontransitive choice. All such models can be formulated in terms of particular functions $v(x,y)$ as in Shafer (1974), defined on the appropriate domains (e.g., lotteries or acts). One important example is *regret theory*, introduced by Loomes and Sugden (1982) (see Starmer, 2000, for a summary), which can be shown to deviate from expected utility only by relaxing transitivity (Diecidue and Somasundaram, 2017). Loomes and Sugden (1987) later extended this framework to *generalized regret theory*, which is based on a function measuring the utility of choosing x net of the regret associated with missing out on y. A second, important example is *salience theory* (Bordalo et al., 2012, 2013), which takes into account the salience of state-dependent payoff differences. Hence, the value attached to a lottery depends on the lottery it is compared to, which allows for nontransitivities. The specific functional valuations proposed by the theory describe a particular function v (defined over lottery pairs) as in Shafer (1974).[2]

Formally, these and other theories allow for nontransitivities because they can be described as special cases of the fundamental representation of Shafer (1974). Specifically, each theory provides a (structural, parametric) functional form for a skew-symmetric function $v(\cdot,\cdot)$ defined on the space which is appropriate for the model, while Shafer (1974) considers an abstract space and an arbitrary function.

2.2 Adding Noise: (Generalized) Random Utility Models

In an *additive* random utility model (McFadden, 1974, 2001, 2005), an agent is assumed to have an underlying utility function u over a feasible set, but to be affected by random utility shocks. Thus, given a choice between two alternatives x and y, realized utilities are $u(x) + \varepsilon_x$ and $u(y) + \varepsilon_y$, respectively, where ε_x, ε_y are mean-zero random variables

[2]Lanzani (2022) proposes a related model of correlation-sensitive choice, which yields formally similar representations and naturally yields nontransitive choices. However, the object of choice are not pairs of lotteries, but rather correlation structures.

(not necessarily independent). Thus, a RUM generates choice probabilities, with the probability of x being chosen when y is also available given by

$$p(x,y) = \text{Prob}(u(x) + \varepsilon_x > u(y) + \varepsilon_y) = \text{Prob}(u(x) - u(y) + \varepsilon_x - \varepsilon_y > 0).$$

where tie-breaking conventions are irrelevant for continuously-distributed errors. Under specific assumptions on the distributions of the error terms, one obtains particular models, as logit choice (Luce, 1959) or probit choice (Thurstone, 1927). This general setting has become one of the dominant approaches in economics to model the fact that choice is empirically (and overwhelmingly) observed to be stochastic.

If the error term $\varepsilon_{xy} = \varepsilon_x - \varepsilon_y$ is assumed to be symmetrically distributed around zero, a preference for x over y is revealed if and only if $p(x,y) \geq 1/2$. Since noise is not directly observable, the assumption of symmetric noise is of course untestable and might be unwarranted. If one is willing to accept it, however, a violation of transitivity in this framework then consists of three (or more) alternatives x, y, z such that $p(x,y) \geq 1/2$, $p(y,z) \geq 1/2$, and $p(z,x) > 1/2$. Hence, a large part of the literature tests for violations of Weak Stochastic Transitivity, which is defined as the condition that if $p(x,y) \geq 1/2$ and $p(y,z) \geq 1/2$, then $p(x,z) \geq 1/2$.

It is important to note, however, that WST fails to properly capture violations of transitivity even in the restricted domain of additive random utility models. It is well-known (Block and Marschak, 1960) that additive random utility models as just described are equivalent to *stochastic preferences*, i.e. probability distributions over transitive preferences. Regenwetter et al. (2010, 2011) and others have argued that violations of transitivity might be due to preferences being unstable in the sense that choices are best described by a probability distribution over transitive preferences, i.e. a stochastic preference. By Block and Marschak (1960) stochastic preferences can also be represented by additive random utility models, as long as error terms are not required to be independent. In particular, such a model (which is included in the class of models we consider) might produce violations of WST even though all involved preferences are transitive, as the following (standard) example shows.

Example 1. There are three alternatives, x, y, and z. A decision maker is described by a distribution over three alternative, transitive preferences: $x \succ y \succ z$, $y \succ z \succ x$, and $z \succ x \succ y$, each with probability 1/3. That is, every time the decision maker makes a choice, one of the three preferences is realized (with equal probabilities) and the decision maker chooses following that preference. In this sense, the decision maker *always* has a transitive preference, which however changes from decision to decision.

It is immediate to see that $p(x,y) = p(y,z) = p(z,x) = 2/3$, and hence WST is violated. Thus WST might be violated even though preferences are described by a standard additive random utility model with transitive preferences. Obviously, however, the corresponding noise terms cannot be symmetric.

The problem of whether a system of choice probabilities can be represented by a stochastic preference (hence an additive random utility model) or not has a well-known solution, with characterizations due to Block and Marschak (1960), Falmagne (1978), McFadden and Richter (1990), and Barberá and Pattanaik (1986). One particularly useful characterization is the *Axiom of Revealed Stochastic Preference* (ARSP; McFadden and Richter, 1990; McFadden, 2005), which states that, for any finite collection of choices $(x_1, y_1), \dots (x_n, y_n)$, one must have that

$$\sum_{i=1}^{n} p(x_i, y_i) \leq \max_{\succ \in \mathcal{P}} \sum_{i=1}^{n} p_{\succ}(x_i, y_i)$$

where $\mathcal{P}$ is the set of all possible strict preferences on the (finite) choice set, and, for any $\succ \in \mathcal{P}$, $p_{\succ}(x, y) = 1$ if $x \succ y$ and $p_{\succ}(x, y) = 0$ if $y \succ x$. That is, the sum of choice probabilities along any sequence of binary choices must be weakly smaller than the largest sum or (degenerate) probabilities one could obtain for a deterministic (transitive) preference. A collection of choice probabilities can be generated by a stochastic preference (or an additive RUM) if and only if it fulfills the ARSP.

Regenwetter et al. (2010, 2011) and others have argued in favor of criteria other than WST to test for stochastic transitivity. In particular, they have proposed to rely on the Triangle Inequality (TI), which (although this fact seems to be largely unmentioned in the literature) is a direct implication of the ARSP (but does not imply it). TI is the condition that, for any three distinct alternatives x, y, z,

$$1 \leq p(x, y) + p(y, z) + p(z, x) \leq 2.$$

The right-hand inequality is immediately implied by the ARSP applied to the choices (x, y), (y, z), (z, x). The left-hand inequality is equivalent to the statement that $p(x, z) \leq p(x, y) + p(y, z)$ (hence the name "Triangle Inequality"), which in turn is equivalent to $p(x, z) + p(z, y) + p(y, x) \leq 2$, which is again just the ARSP applied to the collection of choices (x, z), (z, y), (y, x). Hence, the proposal to use TI is essentially equivalent to testing whether choices can be explained by an additive RUM (although not completely, since the ARSP has implications beyond the TI). The following example (inspired by Birnbaum, 2023) shows that this is also insufficient.

Example 2. There are three alternatives, x, y, and z. A decision maker has a unique, transitive preference: $x \succ y \succ z$. However, the decision maker makes mistakes. Specifically, the decision maker makes a mistake with a 5% probability if confronted with choices (x, y) or (y, z), and with a 25% probability if confronted with the choice (x, z). It follows that $p(x, y) = p(y, z) = 0.95$ and $p(z, x) = 0.25$. this implies that $p(x, y) + p(y, z) + p(z, x) = 2.15 > 2$, and thus the TI (and hence the ARSP) is violated.

This example serves two purposes. First, it exhibits a decision maker who has transitive preferences affected by behavioral noise, but whose choices violate TI. Hence violations of TI are not sufficient to identify genuinely nontransitive preferences in models with noise. Second, since the ARSP is violated, behavior in the above example cannot be represented as an additive RUM. However, the behavior can be represented as arising from a RUM in the sense of Alós-Ferrer et al. (2021), which assumes transitive preferences but allows for pair-specific noise (see below).

Alós-Ferrer et al. (2021) introduced a more general class of RUM models where error terms apply to the utility differences, i.e. the realized utility difference given a choice $\{x, y\}$ is $u(x) - u(y) + \varepsilon_{x,y}$ for a mean-zero random variable $\varepsilon_{x,y}$ and hence

$$p(x, y) = \text{Prob}(u(x) - u(y) + \varepsilon_{x,y} > 0).$$

For instance, Example 2 above can be represented by $u(x) = 3$, $u(y) = 2$, $u(z) = 1$ and continuous random variables ε_{xy}, ε_{yz}, ε_{xz} as follows. For any $v \geq 0$ and any $q \in (0, 1)$, let $\varepsilon^c(-v, q)$ be a continuous random variable with $\text{Prob}(\varepsilon^c(-v, q) \leq -v) = q$ and $E(\varepsilon^c(-v, q)) = 0$.[3] Let ε_{xy} and ε_{yz} be independently distributed as $\varepsilon^c(-1, 0.05)$, and let ε_{xz} be distributed as $\varepsilon^c(-2, 0.25)$. Example 1 can be represented by an additive RUM with correlated error terms, but it is also represented by a RUM in the sense of Alós-Ferrer et al. (2021) with $u(x) = 3$, $u(y) = 2$, $u(z) = 1$ and zero-mean, independent error terms $\varepsilon_{xy} \sim \varepsilon^c(-1, 1/3)$, $\varepsilon_{yz} \sim \varepsilon^c(-1, 1/3)$, and $\varepsilon_{xz} \sim \varepsilon^c(-2, 2/3)$. This illustrates that the class of transitive models that we allow for encompasses arbitrary distributions over transitive preferences plus arbitrary error terms, and is not limited to the classical additive RUMs. Our results will allow us to identify empirical patterns that cannot be generated by *any* transitive model in this class. Those empirical patterns will hence, in particular, exclude that the data is generated by arbitrary RUMs, which also excludes stochastic (unstable) preferences (Regenwetter et al., 2010).

Remark 1. A "trembling-hand model" (e.g. Loomes et al., 2002) assumes that a decision maker is endowed with a fixed (transitive) strict preference but that a pair-specific error might occur. Thus, if $x \succ y$, there is a trembling probability $e_{xy} \in (0, 1)$ that y is chosen. Example 2 is an example of a trembling-hand model.

We claim that any trembling-hand model can be represented as a RUM in the sense of Alós-Ferrer et al. (2021). To see this, consider a trembling-hand model where the preference is represented by a utility function u and the error probabilities $e_{xy} \in (0, 1)$ are as above. For each pair x, y with $x \succ y$ (hence $(u(x) > u(y))$), define a zero-mean continuous random variable ε_{xy} which takes values below $u(y) - u(x)$ with probability e_{xy} (e.g., $\varepsilon_{xy} = \varepsilon^c(u(y) - u(x), e_{xy})$ as defined above). The utility function u together with the noise terms ε_{xy} define a RUM in the sense of Alós-Ferrer et al. (2021).

In "true and error" models (see, e.g., Birnbaum and Schmidt, 2008; Birnbaum, 2023), a decision maker is described by a distribution over preferences (transitive or not) *plus* pair-specific (but preference-independent) error terms. The selected preference is assumed to stay fixed along a given experimental session, and change only across sessions.

[3] For instance, let $\varepsilon^c(-v, q)$ have constant density q on the interval $[-v - 1, -v]$ and constant density $(1 - q)/(K + v)$ on the interval $[-v, K]$, where $K = v + (q/(1 - q))(2v + 1)$.

Hence, for a given session, a true and error model is a trembling-hand model as above, and in particular is encompassed in the class of models we consider.

Remark 2. Random utility models as in Alós-Ferrer et al. (2021) also encompass the class of random parameter models (e.g., Loomes and Sugden, 1998; Apesteguía and Ballester, 2018) as a special case. In those models, a one-parameter functional form for the utility u is fixed. For each choice pair (x, y), a value of the parameter is randomly drawn from a distribution and used to evaluate the choice. This cannot be captured as a standard, additive random utility model, but defines a model in the class considered by Alós-Ferrer et al. (2021). This is because noise in the parameter can be equivalently written as a pair-specific noise term ε_{xy}, which will generally be non-symmetric.

2.3 Random Choice Models and Response Times

We now show how response times can be used to identify genuinely nontransitive preferences. In a framework which assumed transitivity, Alós-Ferrer et al. (2021) provided sufficient conditions on the distributions of response times conditional on each possible choice (x or y for a given pair $\{x, y\}$) which ensure the revelation of a preference for, say, x over y without making any assumptions about the utility function and the distribution of error terms. More precisely, if the conditions are satisfied, the formal results ensure that $u(x) > u(y)$ for any underlying u and any distribution of noise which fits the data (in terms of choices and response times). The importance of these results relies on the fact that they guarantee that an option is preferred to another for *any* utility function and *any* distribution of the error term that the analyst might consider, and hence the results are completely non-parametric and independent of functional forms. The message is that the properties of the empirical distribution of response times allow to recover the underlying preferences in random utility models without imposing any substantive assumptions on the distribution of random terms.

In this subsection, we extend the main result of Alós-Ferrer et al. (2021) to allow for nontransitivities. For this purpose, we go one step forward and consider any skew-symmetric function $v : X^2 \mapsto \mathbb{R}$ (not necessarily arising from a utility function). That is, we consider models where noise is captured by mean-zero random variables $\varepsilon_{x,y}$ and choice probabilities are given by

$$p(x, y) = \text{Prob}(v(x, y) + \varepsilon_{x,y} > 0).$$

We consider abstract options, which could e.g. be themselves lotteries (this will be the case in our empirical analyses). That is, our functions u and v are defined on an abstract space. For a space of lotteries, our approach is agnostic with respect to whether decisions are best represented by expected utility theory, cumulative prospect theory, or any other model of preferences among lotteries. For example, u might be expected utility and v might be any of the functions V^{SSB}, V^R, V^S described in Section 2.1. We merely test the class of models generating transitive choices, where the function above can be written as $v(x, y) = u(x) - u(y)$, against the class of models allowing for nontransitivity lottery choices, where the function $v(x, y)$ cannot be written as a difference of utilities independently of the considered alternatives. The former class includes expected utility theory, rank-dependent utility theory, cumulative prospect theory, and others, while the latter includes generalized regret theory, salience theory, and SSB utility theory.

To derive our result, we need to define what we understand by a dataset. Given the set of alternatives X, denote by $C = \{(x, y) \mid x, y \in X, x \neq y\}$ the set of all binary choice problems, so (x, y) and (y, x) both represent the problem of choice between x and y. Let $D \subseteq C$ be the set of choice problems on which we have data in the form of direct choices, assumed to be non-empty and symmetric, that is, $(x, y) \in D$ implies $(y, x) \in D$. A dataset (including response times) is modeled as follows (Alós-Ferrer et al., 2021).

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

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

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

In an SCF-RT, $p(x, y)$ is interpreted as the frequency with which a decision maker chose x when offered the binary choice between x and y. The assumption that $p(x, y) > 0$ for all $(x, y) \in D$ implies that choice is noisy, that is, every alternative is chosen at least a small fraction of the time. The density $f(x, y)$ describes the distribution of response times conditional on the instances where x was chosen in the binary choice between x and y. The corresponding cumulative distribution function is denoted by $F(x, y)$. The following definition extends the concepts in Alós-Ferrer et al. (2021).

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

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

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

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

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

A RUM-CF is a particular case of RCM-CF where the function v is derived from a utility function, $v(x, y) = u(x) - u(y)$, and hence transitivity is guaranteed. The random variables $\tilde{v}(x, y)$ and their densities $g(x, y)$ capture noisy choice. Condition (RCM.1) requires that noise is unbiased (equivalent to assuming mean zero for an additive term

$\varepsilon_{xy} = \tilde{v}(x,y) - v(x,y)$). Condition (RCM.2) reflects that the choice between x and y is the same as the choice between y and x, and condition (RCM.3) is a regularity condition requiring connected support, i.e. without gaps. Last, r represents the chronometric function, which maps realized values of v into response times $r(|v|)$. Specifically, easier choices (where the value $\tilde{v}(x,y)$ is larger) are faster. This is in keeping with the interpretation that the function v captures a strength of preference.

Given an RCM-CF $(v, \tilde{v}, r)$ and a pair $(x,y) \in C$, the random variable describing the response times predicted by the model conditional on x being chosen over y is given by

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

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

The result we seek will be in terms of preference revelation for *all* RCM-CFs which rationalize (explain) the data. The following definition pins down the formal meaning of the latter.

Definition 3. An RCM-CF $(v, \tilde{v}, r)$ *rationalizes* an SCF-RT (p, f) if

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

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

In other words, an RCM-CF (the model) rationalizes an SCF-RT (the data) if it reproduces both the choice frequencies and the conditional response time distributions in the latter. Obviously, fixing the set D, every RCM-CF predicts an SCF-RT through the equations given in (i) and (ii) above. Thus an alternative definition is that an RCM-CF rationalizes an SCF-RT if the predicted SCF-RT coincides with the actual SCF-RT. We say that an SCF-RT is *rationalizable* if there exists an RCM-CF that rationalizes it. Note that an SCF-RT might be rationalizable by an RCM-CF even though it is not rationalizable by a RUM-CF.

The last definition captures preference revelation in a potentially nontransitive framework.

Definition 4. A rationalizable SCF-RT *reveals that x is preferred to y* if all RCM-CFs that rationalize it satisfy $v(x,y) \geq 0$. It *reveals that x is strictly preferred to y* if all RCM-CFs that rationalize it satisfy $v(x,y) > 0$.

The following Theorem provides our preference revelation result allowing for nontransitive preferences. The proof is in Appendix A.

Theorem 1. *Consider random choice models. A rationalizable SCF-RT (p, f) reveals that x is preferred to y if*

$$p(x,y)F(x,y)(t) \geq p(y,x)F(y,x)(t)$$

for all $t \geq 0$, and that x is strictly preferred to y if the inequality holds for all t and is strict for some t.

The sufficient condition in this result essentially spells out that errors should be slow in a well-defined sense. An intuition is provided in Section 2.4 below. A *Revealed Transitivity Violation* (RTV) exists in the data whenever application of Theorem 1 reveals a preference cycle with $x_1 \succeq x_2 \succeq \ldots \succeq x_n$ and $x_n \succ x_1$. An RTV reveals a nontransitivity which cannot be explained by noise, and thus disentangles noise from genuine transitivity violations. That is, Theorem 1 provides the missing tool that the literature needed to conclude that an apparent transitivity violation does reflect genuinely nontransitive preferences, instead of just being due to behavioral noise. Suppose that a dataset seems to point at nontransitive behavior, e.g. due to a violation of Weak Stochastic Transitivity. That is, the data identify a cycle of, say, three alternatives x, y, z such that $p(x, y) \geq 1/2$, $p(y, z) \geq 1/2$, and $p(z, x) > 1/2$. While a researcher might take this as evidence of a transitivity violation, another researcher might argue that those population frequencies have arisen due to noise (as in a random utility model) even though underlying preferences are transitive. If the dataset includes response times, the researcher can instead apply the "Time Will Tell" (TWT) method derived from Theorem 1 to each of the pairs (x, y), (y, z), and (x, z). If the condition in the Theorem is fulfilled for pairs (x, y) and (y, z), and strictly for (x, z), the researcher can now conclude that nontransitive preferences are revealed, i.e. $x \succeq y$, $y \succeq z$, and $z \succ x$ (an RTV). That is, there exists no random choice model which can explain the data with transitive preferences and behavioral noise, no matter how noise is modeled.

2.4 Intuition for the Revelation Result

Theorem 1, relies on the well-established psychometric and chronometric effects discussed in the introduction. A large literature in psychology, neuroscience, and, more recently, also in economics, has shown that these effects are extremely robust and appear both in perceptual choice (discrimination tasks) and value-based (preferential) choice. They are also standard implications of sequential sampling models from the cognitive sciences as the well-known drift-diffusion model (Ratcliff, 1978; Fudenberg et al., 2018; Webb, 2019; Baldassi et al., 2020).

The psychometric effect is the observation that choices are noisier (and error rates are larger) when alternatives are more similar or, in preference terms, when decision makers are closer to indifference (e.g., Cattell, 1893; Dashiell, 1937; Mosteller and Nogee, 1951; Laming, 1985; Klein, 2001; Wichmann and Hill, 2001; Fudenberg et al., 2018; Alós-Ferrer and Garagnani, 2022a,b). This is implicitly incorporated in standard random utility models, simply because it is more likely that (additive) noise will offset an underlying preference of x over y if the utility difference $u(x) - u(y)$ is small than if it is large. In RCMs, the psychometric effect is captured while allowing for nontransitive choices by replacing the utility difference by the strength of preference $v(x, y)$.

The chronometric effect is the fact that choices are slower when alternatives are more similar or, again in preference terms, when decision makers are closer to indifference (e.g., Cattell, 1902; Dashiell, 1937; Moyer and Landauer, 1967; Moyer and Bayer, 1976; Dehaene et al., 1990; Moffatt, 2005; Chabris et al., 2009; Krajbich et al., 2014, 2015; Fudenberg et al., 2018; Konovalov and Krajbich, 2019; Alós-Ferrer and Garagnani, 2022a,b). This was first incorporated in random utility models in Alós-Ferrer et al. (2021) and is the basis for the revelation result there. In our framework, the chronometric effect is captured by the assumption that the chronometric function r is strictly decreasing in the strength of preference $v(x, y)$.

The intuition for Theorem 1 is as follows. By the chronometric effect, slow choices are more likely to be associated with harder decisions, which in turn result in larger error rates by the psychometric effect. There is, hence, an association between larger noise levels and longer response times. In other words, if an option is chosen systematically more slowly than another one, it is likely that the choice of this option is an error, that is, it contradicts the underlying preference. Essentially, Theorem 1 captures the intuition that, for a preference-based decision, slower choices tend to be errors (see, e.g., Fudenberg et al., 2018, for a discussion of "slow errors").

The difficulty, however, lies in capturing the actual meaning of "slow choices." Since Theorem 1 provides a revelation result which is independent of the model of noise and the underlying function v, "slower" does not correspond to a comparison of means, medians, or any other summary statistic. Rather, the Theorem relies on a sufficient condition which compares the distributions of response times for the two alternatives in a binary choice, namely that

$$p(x, y)F(x, y)(t) \geq p(y, x)F(y, x)(t)$$

for all $t \geq 0$. The left-hand side term in this condition is the probability that x is chosen over y *and* this choice is faster than t. The right-hand side term is the probability that y is chosen over x *and* this choice is faster than t. Thus, the condition requires that, for each given t, x is more likely to be chosen for all decisions that are faster than t. In other words, revealing a preference for x over y requires that y choices are slower than x choices in the sense that, for any fixed deadline t, among the decisions faster than this deadline, there are more x choices than y choices.

Theorem 1 generalizes the main result in Alós-Ferrer et al. (2021), which was restricted to a transitive framework. To see this, rewrite the sufficient condition as

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

which can be seen as a weakening of first-order stochastic dominance, since the condition can only hold if $p(x, y)/p(y, x) \geq 1$. Alós-Ferrer et al. (2021) states this condition as "$F(y, x)$ q-first-order stochastically dominate $F(x, y)$" for $q = p(x, y)/p(y, x)$, which makes clear that the condition balances information from choices and response times. For instance, if choices provide weak information on preferences, $p(x, y)$ and $p(y, x)$ are

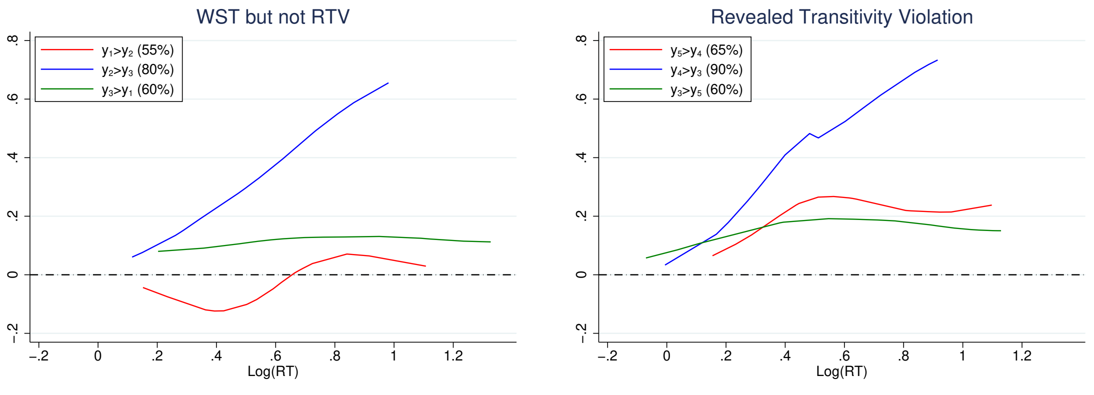

Figure 1: Example 3. Left: A WST violation which does not actually reveal a genuine transitivity violation, i.e. is not an RTV. Right: A WST violation which is revealed to be a genuine transitivity violation (RTV), hence any model of noisy choice rationalizing the data must be nonransitive.

of similar magnitude and the condition approaches first-order stochastic dominance, i.e. it becomes a stringent condition on response times. On the contrary, if choices provide strong evidence on preferences, $p(y, x)$ is close to zero and the condition is increasingly easy to satisfy (becoming void in the limit), i.e. only weak information is required from response times.

2.5 Using Theorem 1

In this subsection, we provide actual examples on how to use Theorem 1 to identify transitivity violations, and how to interpret the results.

Example 3. The first two examples are based on actual data from Kalenscher et al. (2010), whose dataset will be described in detail in Section 3 below. Subjects made repeated binary choices involving the lotteries $y_1 = (\$500, 0.29)$, $y_2 = (\$475, 0.32)$, $y_3 = (\$450, 0.35)$, $y_4 = (\$425, 0.38)$, and $y_5 = (\$400, 0.41)$, which paid the specified amount with the given probability, and zero otherwise.

For subject nr. 26, the choice frequencies for (y_1, y_2), (y_2, y_3), and (y_3, y_1) were

$$p_{26}(y_1, y_2) = 0.55, \quad p_{26}(y_2, y_3) = 0.80, \quad p_{26}(y_3, y_1) = 0.60.$$

For subject nr. 27, the choice frequencies for (y_5, y_4), (y_4, y_3), and (y_3, y_5) were

$$p_{27}(y_5, y_4) = 0.65, \quad p_{27}(y_4, y_3) = 0.90, \quad p_{27}(y_3, y_5) = 0.60.$$

Both patterns are violations of Weak Stochastic Transitivity. Hence, a researcher conducting a traditional analysis would conclude that two violations of transitivity have been found. However, as discussed above (e.g., Example 1), violations of WST can often be accounted for with transitive models, e.g. a distribution over transitive preferences. Hence, a second researcher could argue that the data might be explained by some transitive model and hence constitutes no evidence of transitivity violations.

The experiment of Kalenscher et al. (2010) measured response times, allowing us to obtain the response time distributions $F_{26}(y_1, y_2)$, etc. (see Appendix B for details). Hence, we can use Theorem 1 to ask whether cyclical patterns of preferences are actually revealed. If this is the case, we can conclude that there exists no model explaining the data without violating transitivity. In contrast, if this is not the case, such a conclusion would be unwarranted.

For a preference $y_1 \succ y_2$ to be revealed for subject 26, the condition given in Theorem 1 is that $H_{26}(y_1, y_2)(t) = p_{26}(y_1, y_2)F_{26}(y_1, y_2) - p_{26}(y_2, y_1)F_{26}(y_2, y_1)(t) \geq 0$ for all t, with the inequality being strict for at least some t. In other words, the function $H_{26}(y_1, y_2)$ needs to be above the horizontal axis. Figure 1(left) depicts this function, showing that the condition is indeed fulfilled. We conclude that *any* model (transitive or not) of noisy choice which respects the elementary chronometric effect and rationalizes the data must be such that $y_1 \succ y_2$. Analogously, the figure also shows that $H_{26}(y_2, y_3)$ is above the horizontal axis, and hence a preference $y_2 \succ y_3$ is also revealed. Thus, *any* model of noisy choice rationalizing the data must be such that the underlying preference fulfills $y_1 \succ y_2$ and $y_2 \succ y_3$. If we could also conclude that $y_3 \succ y_1$ is revealed, hence obtaining a nontransitive cycle, this would mean that any model of noisy choice rationalizing the data must be nontransitive. In other words, the conclusion would be that there exists no *transitive* model of noisy choice (respecting the chronometric effect) which can explain the data.

However, as Figure 1(left) shows, the function $H_{26}(y_3, y_1)$ does cross the horizontal axis, hence the condition in Theorem 1 does not hold and the preference $y_3 \succ y_1$ is not revealed. In our analysis in Section 3, whenever we encounter a pattern like this in the data, we classify the apparent violation of transitivity as inconclusive, i.e. a Revealed Transitivity Violation is not obtained.

For subject nr. 27, as Figure 1(right) shows, the analogous function $H_{27}(y_5, y_4)$ is above the horizontal axis, as are the functions $H_{27}(y_4, y_3)(t)$ and $H_{27}(y_3, y_5)(t)$. Applying Theorem 1, we conclude that *any* model of noisy choice which respects the chronometric effect and rationalizes the data must be such that $y_5 \succ y_4$, $y_4 \succ y_3$, and $y_3 \succ y_5$. We have hence identified a genuinely nontransitive cycle. In other words, there exists no transitive model of noisy choice respecting the chronometric effect which rationalizes the data, and no hypothetical researcher will ever be able to argue that some transitive model can account for this data. In our analysis in Section 3, whenever we encounter a pattern like this in the data, we classify it as a genuine violation of transitivity, i.e. a Revealed Transitivity Violation.

Example 4. The following example is a continuous version of Example 1. In that example, a decision maker randomly selects one of three transitive preferences over three alternatives x, y, z. The preferences have a cyclical structure and the resulting choice

Table 1: The Random Choice Model in Example 4.

Prob.	u_k	x	y	z	$v_k(x,y)$	$v_k(y,z)$	$v_k(x,z)$	ε^k_{xy}	ε^k_{yz}	ε^k_{xz}
1/3	u_1	2	1	0	1	1	2	$U[-1,1]$	$U[-1,1]$	$U[-2,2]$
1/3	u_2	0	2	1	-2	1	-1	$U[-2,2]$	$U[-1,1]$	$U[-1,1]$
1/3	u_3	1	0	2	1	-2	-1	$U[-1,1]$	$U[-2,2]$	$U[-1,1]$

probabilities violate WST. The point of that example was that a dataset with these characteristics could not be unambiguously argued to demonstrate a violation of transitivity, because there exists a model involving only transitive preferences which rationalizes the choice probabilities, even though those violate WST.

Technically, that example violates (RCM.3) in Definition 2, because the distribution of utility differences does not have a connected support. Consider, however, the example illustrated in Table 1. The decision maker selects a utility function u_1, u_2, or u_3, each with probability 1/3. The utilities of the alternatives are as given in the table, and the utility differences are also computed there, e.g. $v_1(x,y) = u_1(x) - u_1(y) = 2 - 1 = 1$. However, each utility difference $v_k(a,b)$ is perturbed with a pair-specific noise term ε^k_{ab}, which is uniformly distributed on the interval $[-v_k(a,b), v_k(a,b)]$. This model generates choice probabilities $p(x,y) = p(y,z) = p(z,x) = 2/3$, and hence WST is violated. This is a random choice model, where $v(a,b) = 0$ for each pair and the distribution of utility differences $g(a,b)$ is a convex combination of two uniform distributions centered at 0. Again, this model can be argued to be transitive, since it is just a continuously-perturbed version of a stochastic preference where all involved preferences are transitive. This theoretical construction is of course a knife-edge example, since (because $v(a,b) = 0$ for all three pairs) it corresponds to a RUM with $x \sim y \sim z$.

Suppose, however, that in addition to the choice frequencies, we also observed the conditional, pair-specific distributions of response times. Applying Theorem 1, we can check whether the function $H(a,b)(t) = p(a,b)F(a,b)(t) - p(b,a)F(b,a)(t)$ is always above zero or not, for all three pairs $(a,b) \in \{(x,y),(y,z),(z,x)\}$. Suppose this is the case. Then, a nontransitive cycle $x \succ y \succ z \succ x$ is revealed, which means that there exists no transitive RCM generating the observed choice probabilities and response time distributions. In particular, add an arbitrary (decreasing) chronometric function r to generate response times from the transitive model in Table 1 and. By Theorem 1, the condition on $H(a,b)$ will be violated for at least one of the pairs (a,b), and hence we cannot conclude that the data reveals a genuine transitivity violation.

In other words, a choice dataset with $p(x,y) = p(y,z) = p(z,x) = 2/3$ cannot be argued to demonstrate a violation of transitivity, because there exist (admittedly knife-edge) random choice models based exclusively on transitive preferences that rationalize those choice frequencies. Once the dataset incorporates response times, generating an SCF-RT, the picture changes. There are potential datasets of this type for which the condition in Theorem 1 fails at least once along the potential cycle, and hence it cannot be concluded that transitivity has been violated. In particular, adding an arbitrary chronometric function r to an appropriate transitive model will always generate datasets of this type. There are, however, other potential datasets where (the strict version of) the condition in Theorem 1 is fulfilled for all pairs along the potential cycle. For those datasets, one can conclude that a genuine violation of transitivity has been found. In particular, such a dataset can *never* be generated by adding an arbitrary chronometric function to a transitive choice model.

2.6 Partial Converses to Theorem 1

Theorem 1 provides a sufficient condition for (potentially nontransitive) preference revelation independently of the structure of the noise. For well-established subclasses of models, however, this condition is *necessary and sufficient*, in the sense that if the true data generating process belongs to those classes, the data will *always* fulfill the condition in Theorem 1. Notably, those subclasses include the most-commonly used models in microeconometric analysis and also in psychology. We provide those results below but omit formal definitions of the classes for the sake of brevity.

2.6.1 Symmetric RCMs

First, standard RUMs as used in microeconomics usually assume symmetric noise, e.g. normally distributed or following a Gumbel distribution which generates logit-choice probabilities. Any RUM with Fechnerian noise (Moffatt, 2015) belongs to this class. Generalize this class to include all RCMs, hence allowing for nontransitivity and for pair-specific (symmetric) noise around a strength of preference $v(x, y)$ instead of a utility difference $u(x) - u(y)$, for each pair $(x, y) \in C$. The following result holds.

Proposition 1. *Consider an SCF-RT that is generated by a symmetric RCM-CF. Then, for any $(x, y) \in D$, if $v(x, y) \geq 0$, the condition in Theorem 1 holds, and if $v(x, y) > 0$, the condition is strict for some t.*

The proof is a straightforward adaptation of the proof of Proposition 4 in Alós-Ferrer et al. (2021, p. 1854). Hence, if data is generated by any RUM or RCM where noise is actually symmetric (and any chronometric function), Theorem 1 will always bite, revealing preferences even though the modeler does not assume neither transitivity nor symmetric noise.

2.6.2 The Drift-Diffusion Model

The Drift-Diffusion Model (DDM; e.g., Ratcliff, 1978; Fudenberg et al., 2018) of psychology and neuroscience postulates that binary choice is the result of an internal process of evidence accumulation, formally a diffusion process driven by a drift parameter μ and two (possibly functional) barriers, typically assumed to be either constant or decreasing.

The process randomly drifts up and down, and a choice occurs when the upper or lower barrier is first hit. Which barrier is hit determines the choice, and the time of this event is the response time. This model generates datasets (including response times) which can always be accounted for within our approach.

In particular, let μ_{xy} be the drift rate determining choices for the pair (x, y) through the DDM. Define $v(x, y) = \mu_{xy}$ for each (x, y). This defines an RCM-CF, with potentially nontransitive choices. The following result holds.

Proposition 2. *Consider an SCF-RT that is generated by a DDM with constant or decreasing boundaries, without any constraints on the underlying drift rates μ_{xy}. Then, for any $(x, y) \in D$, if $\mu_{xy} \geq 0$, the condition in Theorem 1 holds, and if $\mu_{xy} > 0$, the condition is strict for some t.*

The proof is a direct adaptation of the proof of Proposition 6 in Alós-Ferrer et al. (2021, p. 1858). Thus, if the DDM is the true data generating process, the resulting dataset always fulfills the conditions we identify in Theorem 1 below, and hence the sign of $v(x, y)$ is always revealed, even though the modeler does not assume that the DDM is the correct model and transitivity might not hold.

Remark 3. Given the last observation, instead of considering a general class of models, we could have particularized our entire approach to the analysis of the DDM as a specific decision-making process, and then our results would reveal whether data can be explained through the DDM with a transitive structure, e.g. when $\mu_{xy} = u(x) - u(y)$ for some utility function u or not. We refrain from this approach because then it could be potentially argued that the violations of transitivity that we find might be explainable by a transitive model based on a more general process model than the DDM. Our approach does not rely on a specific model and hence escapes that potential critique.

3 Empirical Evidence for Nontransitivity

3.1 Description of the Datasets

In this section we apply Theorem 1 to two existing datasets, both of which were specifically collected to study transitivity violations. The selected datasets, from Davis-Stober et al. (2015) (DSBC) and Kalenscher et al. (2010) (KTHDP), are ideal for our purposes because they include response times and every participant repeated every choice a reasonable number of times.

In the dataset of DSBC, $N = 60$ subjects made binary choices among different lotteries in a 2×2 within-subject design. Specifically, the experiment varied the display format of the lotteries (pies vs. bars) and whether participants faced a time constraint when making their choices or not (4 seconds vs. no time limit). The choice pairs were drawn from two sets of five lotteries each, with one lottery common to the two sets. All possible combinations of the lotteries within each set were implemented, giving rise to 20

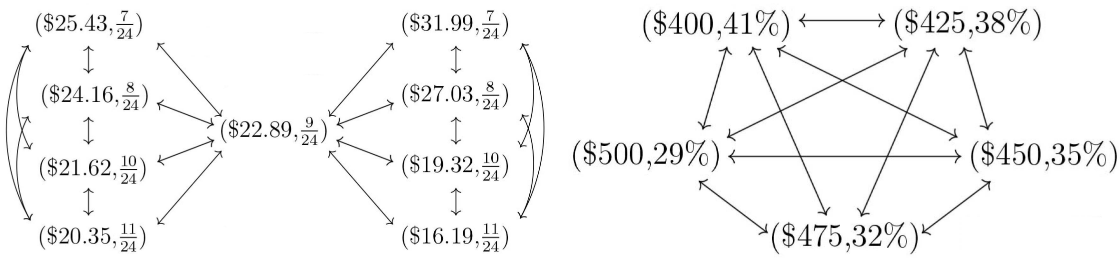

Figure 2: List of lotteries and implemented pairwise comparisons in Davis-Stober et al. (2015) (left) and Kalenscher et al. (2010) (right).

distinct choice pairs (see Figure 2, left). Each of these pairs was repeated 12 times in each of the 4 possible conditions, for a total of $12 \times 4 \times 20 = 960$ choices per participant. Each participant took part in four sessions (in two non-consecutive days), with two (randomly allocated) combinations of time limits and display format manipulations in each of them. Sessions lasted an average of 45 minutes (multi-session experiments allowing for a large number of choices and standard-length sessions are common; e.g., Hey, 2001). Choices were incentivized (one decision from each condition was randomly selected and paid, in addition to a show-up fee).

In the dataset of KTHDP, $N = 30$ subjects made binary choices among five different lotteries.[4] All lottery combinations were implemented (see Figure 2, right). Each of the 10 resulting choice pairs was repeated 20 times, for a total of 200 trials per participant. There was a time limit of 4 seconds, with misses resulting in missed trials. Each participant took part in a single, individual-level session while being scanned in an fMRI machine. Sessions lasted an average of 49 minutes. Choices were incentivized (with dummy dollars translated into Euro with a conversion rate of 100:1), with one randomly-selected decision paid in addition to a show-up fee.

In addition to the presence of repetitions, the measurement of response times, and the fact that they were collected to study transitivity violations, the two datasets are also interesting for other reasons. First, all lotteries involve only one non-zero outcome and hence can be presented with only two variables (a single outcome and its probability). This makes alternatives easy to compare for participants. Second, all magnitudes in each of the experiments are comparable (without extreme differences), hence mitigating possible concerns regarding range or outlier effects. Third, none of the lotteries involves probabilities close to zero or one, which are known to generate their own regularities.

3.2 Revealed Transitivity Violations

We now investigate Revealed Transitivity Violations in the two datasets. That is, we apply Theorem 1 to the two datasets to reveal preference cycles with $x_1 \succsim x_2 \succsim \ldots \succsim x_n$ and $x_n \succ x_1$. An RTV reveals a nontransitivity which cannot be explained by noise, and

[4] Further 240 filler lotteries were used, but they all were paired in a way which involved dominated choices, and hence are not interesting for our purposes.

thus disentangles noise from genuine transitivity violations. Since, within the universe of RCMs, preferences revealed by our method are independent of any assumptions (i) about noise and (ii) about functional forms regarding $v(x, y)$, we conclude that transitivity violations identified by our method cannot be due to any form of noise or functional form assumption regarding v.

If subjects had transitive preferences, empirically-observed violations of the conditions previously used in the literature, e.g. Weak Stochastic Transitivity, would be due to noise. Then, once we identify the set of cycles of alternatives for which choices reveal preferences, the subset of RTVs should be empty. On the other hand, if transitivity violations arise from a genuinely nontransitive preference, then the subset of cycles of alternatives where all preferences are revealed should still contain violations of transitivity, i.e. RTVs.

We start by applying our method to reveal preferences. That is, for every (potential) cycle of alternatives in the design

$$(x_1, x_2, \ldots, x_n, x_{n+1} = x_1)$$

and every pair (x_i, x_{i+1}) of subsequent alternatives along the cycle, and for every participant, we compute the actual choice proportions across repetitions and the response time densities in the data, and check whether the condition in Theorem 1 holds (see Appendix B for further details on the empirical procedure). For DSBC, the average percentage of choices at the subject level for which the method reveals preferences is 56.67% (median 57.22%, SD= 6.15, min 44.66%, max 68.31%), while for KTHDP is 77.00% (median 75.00%, SD=15.57, min 40.00%, max 100.00%). Thus, in our datasets, the method reveals preferences often enough for an analysis of revealed nontransitivities to be conducted.[5]

Say that a cycle of alternatives $(x_1, x_2, \ldots, x_n, x_{n+1} = x_1)$ is a *revealed cycle* if all preferences along the cycle are revealed, i.e. for every pair (x_i, x_{i+1}) of subsequent alternatives along the cycle, $i = 1, \ldots, n$, the method reveals either $x_i \succeq x_{i+1}$ or $x_{i+1} \succeq x_i$ (or the corresponding strict preferences). For example, a cycle of alternatives (x_1, x_2, x_3, x_1) could be a revealed cycle if the method revealed $x_1 \succeq x_2$, $x_2 \succeq x_3$ and $x_1 \succeq x_3$ (which is compatible with transitivity), but it would also be a revealed cycle if the method revealed $x_1 \succeq x_2$, $x_2 \succeq x_3$ and $x_3 \succ x_1$ (which violates transitivity and hence is an RTV).

The proportion of revealed cycles is obviously smaller than the proportion of choices for which preferences are revealed, since all preferences along a cycle of alternatives must be revealed for the cycle to be revealed. For DSBC, 20.82% of cycles of alternatives are

[5]For DSBC participants, we find no differences in the proportion of revealed preferences depending on whether subjects had time limits or not (56.13% vs. 57.16%; WRS, $N = 60$, $z = -0.942$, $p = 0.3505$). This is important, as it suggests that even though the method relies on response times, its capacity to reveal preferences is not affected by (reasonable) time limits, and hence it is robust with respect to such manipulations. In the Online Appendix we take advantage of the manipulations in DSBC (time limits and graphical formats) to further investigate the robustness of the results.

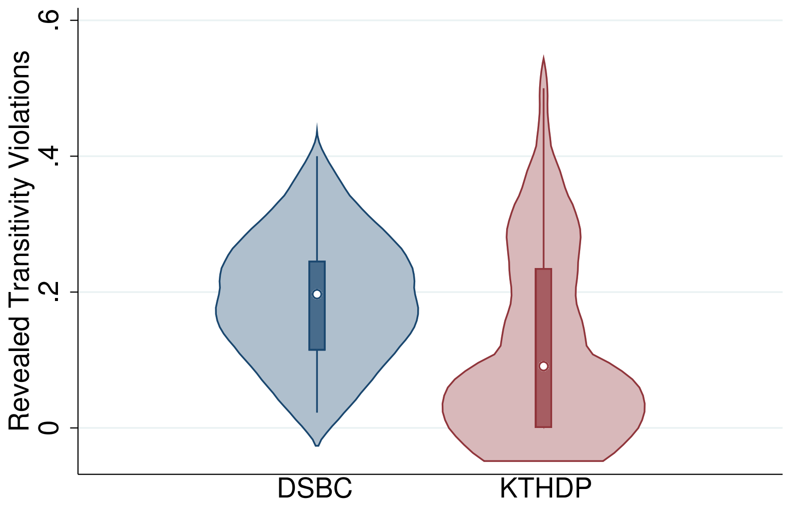

Figure 3: Distribution of the individual proportions of RTVs over all cycles of alternatives where all preferences are revealed. Violin plots show the median, the interquartile range and the 95% confidence intervals as well as rotated kernel density plots on each side. Fifty percent of individuals exhibit 19.69% (9.09%) or more Revealed Transitivity Violations in the DSBC data (KTHDP data).

revealed (median 22.08%, SD=7.61, min 0.00%, max 32.08%), and the number is 54.25% (median 60.00%, SD=25.00, min 0.00%, max 100.00%) for KTHDP.

For each individual in each of the two datasets, we identified the set of revealed cycles of alternatives and then checked for which of those the revealed preferences were nontransitive. We found sizeable sets of Revealed Transitivity Violations. The average individual proportion of RTVs, that is, the proportion of all revealed cycles of alternatives which are RTVs, is 19.24% (SD=8.48) in DSBC and 13.83% (SD=15.41) in KTHDP. Figure 3 plots the distribution of subject-level proportions of RTV over all revealed cycles of alternatives for both datasets, revealing considerable heterogeneity. For DSBC, the individual proportion of RTVs ranges from 2.25% to 40.00%, with a median of 19.69%. For KTHDP, the individual proportion of RTVs ranges from 0.00% to 50.00%, with a median of 9.09%. This also implies that Revealed Transitivity Violations are pervasive in the two datasets, since fifty percent of individuals exhibit 19.69% (9.09%) or more RTVs in the DSBC data (KTHDP data).

In summary, our approach identifies transitivity violations which *cannot* be explained by noise (at least within the framework of RCMs), and hence the set of violations we identify stand on conceptually solid ground as a demonstration that nontransitivities in the data do occur.

We remark that the data also contains cycles of alternatives for which not all preferences along the cycle are revealed, for which we hence cannot unambiguously determine whether preferences are nontransitive. The size of the set of RTVs we identify is hence a lower bound on genuine transitivity violations. The important realization is that this lower bound is not zero, that is, genuinely nontransitive preferences are present in both datasets.

3.3 Limitations of the datasets

Our empirical analysis is based on the two datasets of DSBC and KTHDP. These datasets are ideal for our purposes because they provide a large number of decisions per individual and choice pair and also contain response times. Below we address some possible limitations and criticisms.

3.3.1 Inattention

Since both datasets involve a large number of choices per participant, one might worry whether part of the choice errors might arise from distraction or inattention and interact with our conclusions. This is not the case.

First, as reported in Section 3.1, sessions in both experiments were within standard limits in experimental economics and psychology. Sessions in DSBC lasted an average of 45 minutes, with each participant taking part in four sessions. In the fMRI experiment of KTHDP, instructions, questionnaires, and payment were all administered outside of the scanner, and the actual decision-making session lasted an average of 49 minutes. Hence, data collection was well within standards in both cases.

Second, we can test for possible effects of subject-level attention by splitting the data of each participant according to whether it was collected in the first or the second half of the corresponding session. There are no statistically significant differences in (i) the individual proportion of pairs where preferences are revealed according to Theorem 1, (ii) the individual proportion of Revealed Transitivity Violations, or (iii) the individual proportion of WST violations. Additionally, the dataset of KTHDP includes a number of choice pairs where one option was first-order stochastically dominated. Violations of stochastic dominance could be an indicator of individual-level inattention. However, we found no significant correlation between the individual proportion of dominated choices and the individual proportions (i), (ii), and (iii) listed above. Also, the individual proportions of dominated choices in this dataset are not statistically significantly different between the first and the second half of the session. All tests are given in Online Appendix C.

Third, the literature on the psychology of attention shows that inattention typically results in *longer* response times (Wilding, 1971; Luce, 1986; Novikov et al., 2017; De Boeck and Jeon, 2019). Hence, if part of the intransitivities we observe are simply errors due to distraction, this would make the associated response times longer. This is hence just another version of the intuitive argument provided in the introduction: if received transitivity violations were just due to noise, they should be associated with slow choices. However, the analysis mentioned in the introduction finds precisely the opposite (a substantial number of transitivity violations where all involved choices are fast). As discussed in Section 2.4, the preference revelation result in Theorem 1 formalizes the actual concept of "slow errors," and hence possible distraction leading to more and slower errors is not a problem for the approach.

Last, it is of course possible in any given experiment that some extraneous source of noise, orthogonal to payoffs, affects a particular set of trials, generating data which does not reflect underlying preferences. The condition in Theorem 1, however, is a *sufficient* condition which guarantees preference revelation independently of the noise. If choices are truly unrelated to preferences due to extraneous noise, the outcome in practice will be a failure of the sufficient condition. That is, a noisy experiment will simply result in a low proportion of revealed preferences. The point of our analysis above is that the sufficient condition holds often enough to identify a significant amount of transitivity violations which cannot be explained by any model of noise.

3.3.2 Time limits

The experiment of KTHDP, and some sessions in DSBC, involved time limits for the decisions. This is a standard practice in experiments with repeated decisions to control session length. However, it does not qualify as time pressure unless a large part of the decisions are actually constrained by the limits (which is usually ensured with appropriate pre-measurements if time pressure is to be used). In the two experiments we consider, the imposed time limits were typically non-binding. For DSBC the average response time for the condition without time limits was 2.010 s, while the time limit in the other condition was 4 s, and only 1.10% of observations failed to meet this deadline. For KTHDP, all choices were under the same 4 s time limit, the average response time was 1.484 s, and only 0.59% of observations missed the deadline. Hence, there was no actual time pressure in any of the experiments.

Since DSBC included sessions with and without time limits, we can directly test whether those had any effect on our variables of interest. This is not the case. There were no significant differences between sessions with and without time limits for the individual proportion of pairs where preferences are revealed according to Theorem 1, the individual proportion of Revealed Transitivity Violations, or the individual proportion of WST violations. All tests are given in Online Appendix D.

4 Comparing RTVs and WST

Our results in the previous section show that, when cycles of alternatives are revealed, genuinely nontransitive preferences exist in a substantial number of cases. As just mentioned, there are also cycles of alternatives where the sufficient condition in Theorem 1 fails for at least one choice, and hence preferences are not revealed. However, there appears to be little reason to believe that the share of genuine nontransitivities should be different for those cycles of alternatives for which not all preferences are revealed.

As an indication of this, in this section we show that the percentage of RTVs over all revealed cycles of alternatives is not significantly different from the percentage of WST violations among all cycles of alternatives.

4.1 Relation of RTVs to Weak Stochastic Transitivity

Up to now, the empirical literature has predominantly looked at violations of *Weak Stochastic Transitivity* (WST) to study transitivity violations. This property states that for all x_1, x_2, x_3 such that $p(x_1, x_2) \geq 1/2$ and $p(x_2, x_3) \geq 1/2$, it must follow that $p(x_1, x_3) \geq 1/2$. Other concepts of transitivity in a stochastic setting exist, as e.g. strong stochastic transitivity (where the implication is that $p(x_1, x_3) \geq \max\{p(x_1, x_2), p(x_2, x_3)\}$), moderate stochastic transitivity (which replaces the maximum with the minimum in the previous implication; see He and Natenzon, 2022), or the Triangle Inequality (recall Section 2). See Fishburn (1998) for an overview. However, WST remains a natural choice as a benchmark given our theoretical framework, and we will use it for ease of comparison to the literature. Further, since strong stochastic transitivity implies moderate stochastic transitivity and the latter implies WST, every violation of WST implies violations of the former two properties.

There is a relation between RTVs and violations of WST, which is derived from the following observation. Note that the condition in Theorem 1 implies that $p(x, y) \geq p(y, x)$ (e.g., by taking limits as $t \to \infty$) even if this were not stated as part of the definition. That is, if Theorem 1 reveals a (nontransitive) preference for x over y, if follows that $p(x, y) \geq 1/2$. Conversely, if $p(x, y) > 1/2$ and a preference between these two alternatives is revealed, only a preference of x over y can be revealed. That is, preferences cannot be revealed "against" choice frequencies, but choice frequencies do not imply preference revelation. This property delivers a link to Weak Stochastic Transitivity, because the latter is stated in terms of choice frequencies.

First, it follows from this property that an RTV implies a violation of WST, except in knife-edge cases. That is, the concept of RTV is in practice more stringent than violations of WST. The argument is as follows. If a nontransitive preference cycle $x_1 \succeq x_2 \succeq x_3 \ldots x_n \succ x_{n+1} = x_1$ is revealed by Theorem 1, it follows from the comment above that $p(x_i, x_{i+1}) \geq 1/2$ for all $i = 1, \ldots, n$. Hence, this cycle must also entail a WST violation, except in the knife-edge case where $p(x_1, x_2) = p(x_2, x_3) = \ldots = p(x_n, x_1) = 1/2$. That is, every RTV where not all choice frequencies are 1/2 is also a WST violation, meaning that the concepts are in practice naturally nested.

In principle, however, some WST violations might not be RTVs, and we are interested in the proportion of empirical WST violations which are RTVs, and for which the researcher is thus actually justified to infer the existence of genuinely nontransitive preferences. Suppose a sequence of alternatives $x_1, x_2, \ldots, x_n, x_{n+1} = x_1$ builds a violation of Weak Stochastic Transitivity, i.e. $p(x_i, x_{i+1}) \geq 1/2$ for all $i = 1, \ldots, n-1$ and $p(x_n, x_1) > 1/2$. The researcher then applies Theorem 1 to the pairs along the

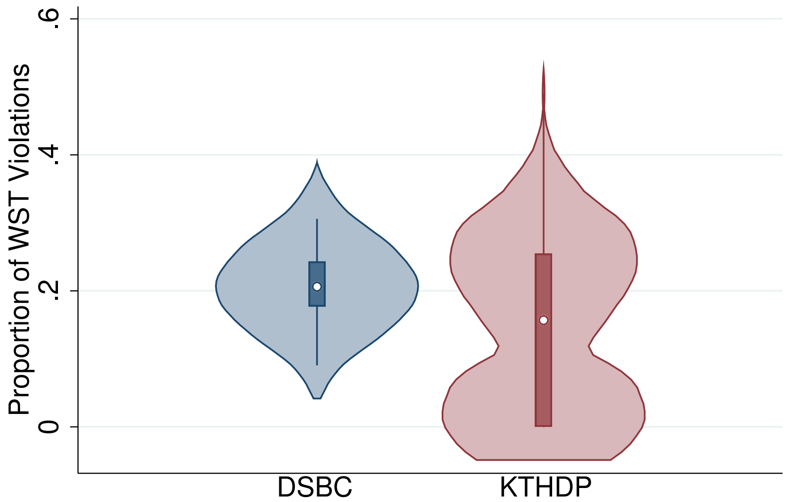

Figure 4: Distributions of the individual proportions of Weak Stochastic Transitivity violations, computed over *all* cycles of alternatives in the datasets. Violin plots show the median, the interquartile range and the 95% confidence intervals as well as rotated kernel density plots on each side. Fifty percent of individuals exhibit 20.61% (15.69%) or more WST violations in the DSBC data (KTHDP data).

sequence. In view of the comment above, essentially only two outcomes are possible. In the first case, preferences fail to be revealed for at least one of the pairs. In this case, the researcher is not entitled to conclude that the observed violation of Weak Stochastic Transitivity is actually due to a nontransitivity in underlying preferences; in other words, the observed violation might well be due to noise.

In the second case, preferences are revealed for all pairs along the cycle. In practice (and in our data), this case always implies an RTV. That is, every WST violation for which all preferences along the cycle of alternatives are revealed is actually an RTV except for knife-edge cases. Specifically, suppose the choice frequencies do not involve any exact tie, $p(x_i, x_{i+1}) > 1/2$ for all $i = 1, \ldots, n$. By the comment above, since preferences cannot be revealed against choice frequencies, revealed preferences necessarily must form a nontransitive preference cycle, i.e. an RTV, except in the knife-edge case of revealed full indifference.

If a WST violation involves an exact tie in choice frequencies, it cannot be concluded that an RTV will be obtained, even if all preferences are revealed. This is because if $p(x, y) = 1/2$, the TWT method might reveal a strict preference either way, hence allowing for WST violations involving frequency ties where all preferences are (even strictly) revealed but a nontransitive cycle does not arise. In practice, such knife-edge cases are empirically rare and they never occurred in our data.

4.2 Empirical Comparison Between RTVs and WST Violations

We now compare RTVs to WST violations, as studied in the previous literature on transitivity. Figure 4 displays violin plots for the subject-level proportion of WST violations computed over all cycles of alternatives, in both datasets. For DSBC, we observe that, on average across individuals, 20.77% of all cycles of alternatives in the dataset result in WST violations (median 20.61%, SD=5.28, min 9.04%, max 34.57%), while in KTHDP the average is 15.42% (median 15.69%, SD=13.93, min 0.00%, max 49.02%). These proportions are roughly representative of results in the literature, and indicate a sizeable percentage of transitivity violations if WST is used as a criterion.

Recall that the concept of RTV is in practice more stringent than violations of WST (Section 4.1). We are hence interested in the proportion of empirical WST violations which are RTVs, and for which the researcher is thus actually justified to infer the existence of genuinely nontransitive preferences, i.e., for which the nontransitivity cannot be explained by any model of noise. To compare revealed nontransitivities according to Theorem 1 with violations of WST, we first compute the proportion of all WST violations that are actually RTVs. In practice, those coincide with the WST violations where the cycle of alternatives is revealed (again, recall Section 4.1). We obtain that, on average across subjects, 19.24% of all WST violations are actually RTVs for DSBC (median 17.71%, SD=9.56, min 4.35%, max 43.42%). The average is 39.58% for KTHDP (median 29.41%, SD=32.60, min 0.00%, max 100.00%). This means that for 19.24% of all WST violations for DSBC, and 39.58% for KTHDP, application of Theorem 1 reveals transitivity violations that uncover genuinely nontransitive preferences and that cannot be due to noise. For the remaining (non-RTV) observed WST violations, it cannot be discarded that they may be due to some sort of underlying noise, but it also cannot be discarded that they may be due to genuinely nontransitive preferences.

We would also like to quantify the *size* of the set of transitivity violations at the individual level, and compare it to previous measurements using WST. Since the number of RTVs for a given subject is necessarily smaller than the individual number of WST violations (Section 4.1), a direct comparison would just mechanically show that there are less RTVs than WST violations. Thus, we compare the proportions relative to the relevant sets in each case. That is, we compare the proportion of RTVs in relation to cycles of alternatives with revealed preferences only (as discussed in Section 3.2 and illustrated in Figure 3) with the proportion of WST violations in relation to all cycles of alternatives, revealed or not (as illustrated in Figure 4). These proportions are not mechanically related to each other, and hence this procedure allows a fair comparison of the magnitudes of transitivity violations as suggested by RTV and WST.

If violations of WST would mainly arise from choices which are not revealed, we should see a sharp decrease in the proportion of transitivity violations according to RTV when computed in this way (since non-revealed cycles of alternatives are excluded), when compared to WST violations. On the contrary, if violations of WST are orthogonal to whether preferences are revealed by Theorem 1 or not, the overall proportion of transitivity violations according to WST and to RTV should be unaffected.

Recall that the individual proportion of RTVs in DSBC was 19.24%, compared to a proportion of 20.77% of WST violations for the overall sample. The difference is small, and a Wilcoxon Rank-Sum test reveals no significant differences at the 5% level ($N = 60$, $z = -1.811$, $p = 0.0705$). In KTHDP the proportion of RTVs was 13.83%, compared to a 15.42% of WST violations for the overall sample. Again there are no significant differences at the 5% level (WRS, $N = 29$, $z = -1.847$, $p = 0.0657$). Hence, the evidence is aligned with the interpretation that transitivity violations might be orthogonal to whether preferences are revealed by Theorem 1 or not. However, of course, this is just suggestive evidence and one cannot conclude that WST violations where preferences are not revealed are actually transitivity violations.

5 Discussion

Are economic choices transitive? A long-standing discussion in economics has addressed this fundamental issue. A negative answer would shake the very foundations of applied microeconomic analysis, and empirical evidence to this effect has been, understandably, subjected to detailed scrutiny. In particular, evidence in favor of transitivity violations has been systematically criticized as deriving from behavioral noise.

In this paper we provide a new formal framework which allows to reveal "preferences" even when they are not transitive, disentangling them from behavioral noise. We then derive a non-parametric method which generalizes recent preference revelation results using both choice frequencies and response times, but allows for nontransitive preferences. We apply this method to two distinct datasets and find conclusive evidence that, even when one fully disentangles behavioral noise from underlying preferences, transitivity violations are reduced but do not disappear. In this sense, transitivity violations are not a mere artefact of the analysis or a consequence of behavioral noise, but rather an actual feature of human behavior.

We view our results as a call for attention. The fundamental assumption that economic choices can be explained by transitive preferences is useful but wrong, in the sense that it does not always hold even if one allows for behavioral noise. Any model that assumes that people evaluate alternatives independently of other alternatives and tend to choose the option with the higher overall evaluation satisfies transitivity, and hence stands on somewhat-shaky grounds. This includes of course normative models as expected utility theory, but also descriptive models built to accommodate behavioral anomalies as cumulative prospect theory (Tversky and Kahneman, 1992) and many others. Ultimately, applied economics needs to embrace models allowing for violations of transitivity. Those are still sparse (e.g. Shafer, 1974; Loomes and Sugden, 1982; Fishburn, 1982, 1986; Bordalo et al., 2012; Lanzani, 2022), but include some prominent examples as salience theory and regret theory.

References

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

Alós-Ferrer, C. and M. Garagnani (2022a). Strength of Preference and Decisions Under Risk. *Journal of Risk and Uncertainty 64*(3), 309–329.

Alós-Ferrer, C. and M. Garagnani (2022b). The Gradual Nature of Economic Errors. *Journal of Economic Behavior and Organization 200*, 55–66.

Anderson, S. P., J.-F. Thisse, and A. De Palma (1992). *Discrete Choice Theory of Product Differentiation*. Cambridge, MA: MIT Press.

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

Baldassi, C., S. Cerreia-Vioglio, F. Maccheroni, and M. Marinacci (2020). A Behavioral Characterization of the Drift Diffusion Model and its Multi-Alternative Extension to Choice under Time Pressure. *Management Science 66*(11), 5075–5093.

Barberá, S. and P. K. Pattanaik (1986). Falmagne and the Rationalizability of Stochastic Choices in Terms of Random Orderings. *Econometrica 54*(3), 707–715.

Birnbaum, M. H. (2020). Reanalysis of Butler and Pogrebna (2018) Using True and Error Model. *Judgment and Decision Making 15*(6), 1044–1051.

Birnbaum, M. H. (2023). Testing Transitivity of Preference in Individuals. *Decision 10*(2), 153–180.

Birnbaum, M. H. and U. Schmidt (2008). An Experimental Investigation of Violations of Transitivity in Choice Under Uncertainty. *Journal of Risk and Uncertainty 37*(1), 77–91.

Birnbaum, M. H. and U. Schmidt (2010). Testing Transitivity in Choice Under Risk. *Theory and Decision 69*(4), 599–614.

Bleichrodt, H. and P. P. Wakker (2015). Regret theory: A bold alternative to the alternatives. *Economic Journal 125*(583), 493–532.

Block, H. D. and J. Marschak (1960). Random Orderings and Stochastic Theories of Responses. In I. Olkin (Ed.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, pp. 97–132. Stanford: Stanford University Press.

Bordalo, P., N. Gennaioli, and A. Shleifer (2012). Salience Theory of Choice under Risk. *Quarterly Journal of Economics 127*(3), 1243–1285.

Bordalo, P., N. Gennaioli, and A. Shleifer (2013). Salience and Consumer Choice. *Journal of Political Economy 121*(5), 803–843.

Butler, D. and G. Pogrebna (2018). Predictably Intransitive Preferences. *Judgment and Decision Making 13*(3), 217–236.

Cattell, J. M. (1893). On Errors of Observation. *The American Journal of Psychology 5*(3), 285–293.

Cattell, J. M. (1902). The Time of Perception as a Measure of Differences in Intensity. *Philosophische Studien 19*, 63–68.

Chabris, C. F., C. L. Morris, D. Taubinsky, D. Laibson, and J. P. Schuldt (2009). The Allocation of Time in Decision-Making. *Journal of the European Economic Association 7*(2-3), 628–637.

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

Davis-Stober, C. P., N. Brown, and D. R. Cavagnaro (2015). Individual Differences in the Algebraic Structure of Preferences. *Journal of Mathematical Psychology 66*, 70–82.

De Boeck, P. and M. Jeon (2019). An Overview of Models for Response Times and Processes in Cognitive Tests. *Frontiers in Psychology 10*, 102.

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

Diecidue, E. and J. Somasundaram (2017). Regret Theory: A New Foundation. *Journal of Economic Theory 172*, 88–119.

Dyer, J. S. and R. K. Sarin (1982). Relative Risk Aversion. *Management Science 28*(8), 875–886.

Falmagne, J.-C. (1978). A Representation Theorem for Finite Random Scale Systems. *Journal of Mathematical Psychology 18*(1), 52–72.

Fishburn, P. C. (1982). Nontransitive Measurable Utility. *Journal of Mathematical Psychology 26*(1), 31–67.

Fishburn, P. C. (1984a). Dominance in SSB Utility Theory. *Journal of Economic Theory 34*(1), 130–148.

Fishburn, P. C. (1984b). SSB utility theory: An economic perspective. *Mathematical Social Sciences 8*(1), 63–94.

Fishburn, P. C. (1984c). SSB Utility Theory and Decision-Making Under Uncertainty. *Mathematical Social Sciences 8*(3), 253–285.

Fishburn, P. C. (1986). Ordered Preference Differences Without Ordered Preferences. *Synthese 67*(2), 361–368.

Fishburn, P. C. (1988). *Nonlinear Preference and Utility Theory.* Baltimore, Maryland: Johns Hopkins University Press.

Fishburn, P. C. (1998). Stochastic Utility. In S. Barberà, P. J. Hammond, and C. Seidl (Eds.), *Handbook of Utility Theory*, Volume 1: Principles, Chapter 7, pp. 273–319. Kluwer Academic Publishers.

Fudenberg, D., P. Strack, and T. Strzalecki (2018). Speed, Accuracy, and the Optimal Timing of Choices. *American Economic Review 108*(12), 3651–3684.

He, J. and P. Natenzon (2022). Moderate Expected Utility. Working Paper, Washington University in Saint Louis, available at SSRN 3243657.

Hey, J. D. (2001). Does Repetition Improve Consistency? *Experimental Economics 4*(1), 5–54.

Humphrey, S. (2001). Non-transitive Choice: Event-Splitting Effects or Framing Effects? *Economica 68*(269), 77–96.

Iverson, G. and J.-C. Falmagne (1985). Statistical Issues in Measurement. *Mathematical Social Sciences 10*(2), 131–153.

Kalenscher, T., P. N. Tobler, W. Huijbers, S. M. Daselaar, and C. Pennartz (2010). Neural Signatures of Intransitive Preferences. *Frontiers in Human Neuroscience 4*(49), 1–14.

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

Konovalov, A. and I. Krajbich (2019). Revealed Strength of Preference: Inference from Response Times. *Judgment and Decision Making 14*(4), 381–394.

Krajbich, I., B. Bartling, T. Hare, and E. Fehr (2015). Rethinking Fast and Slow Based on a Critique of Reaction-Time Reverse Inference. *Nature Communications 6*(7455), 1–9.

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

Kreweras, G. (1961). Sur une possibilité de rationaliser les intransitivités. In *La Décision, Colloques Internationaux du Centre National de la Recherche Scientifique*, pp. 27–32. Paris: Editions du Centre National de la Recherche Scientifique.

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

Lanzani, G. (2022). Correlation Made Simple: Applications to Salience and Regret Theory. *Quarterly Journal of Economics 137*(2), 959–987.

Lo, A. (2019). Demystifying the Integrated Tail Probability Expectation Formula. *The American Statistician 73*(4), 367–374.

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

Loomes, G., C. Starmer, and R. Sugden (1989). Preference Reversal: Information-Processing Effect or Rational Non-Transitive Choice? *Economic Journal 99*(395), 140–151.

Loomes, G., C. Starmer, and R. Sugden (1991). Observing Violations of Transitivity by Experimental Methods. *Econometrica 59*(2), 425–439.

Loomes, G. and R. Sugden (1982). Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty. *Economic Journal 92*(368), 805–824.

Loomes, G. and R. Sugden (1987). Some Implications of a More General Form of Regret Theory. *Journal of Economic Theory 41*(2), 270–287.

Loomes, G. and R. Sugden (1998). Testing Different Stochastic Specifications of Risky Choice. *Economica 65*(260), 581–598.

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

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

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

McFadden, D. L. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*, pp. 105–142. New York: Academic Press.

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

McFadden, D. L. (2005). Revealed Stochastic Preference: A Synthesis. *Economic Theory 26*(2), 245–264.

McFadden, D. L. and M. K. Richter (1990). Stochastic Rationality and Revealed Preference. In J. S. Chipman, D. L. McFadden, and M. K. Richter (Eds.), *Preferences, Uncertainty, and Optimality: Essays in Honor of Leonid Hurwicz*, pp. 163–186. Boulder, Colorado: Westview Press.

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

Moffatt, P. G. (2015). *Experimetrics: Econometrics for Experimental Economics.* London: Palgrave Macmillan.

Mosteller, F. and P. Nogee (1951). An Experimental Measurement of Utility. *Journal of Political Economy 59*, 371–404.

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

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

Novikov, N. A., Y. M. Nurislamova, N. A. Zhozhikashvili, E. E. Kalenkovich, A. A. Lapina, and B. V. Chernyshev (2017). Slow and Fast Responses: Two Mechanisms of Trial Outcome Processing Revealed by EEG Oscillations. *Frontiers in Human Neuroscience 11*, 218.

Ranyard, R., H. Montgomery, E. Konstantinidis, and A. L. Taylor (2020). Intransitivity and Transitivity of Preferences: Dimensional Processing in Decision Making. *Decision 7*(4), 287–313.

Ratcliff, R. (1978). A Theory of Memory Retrieval. *Psychological Review 85*, 59–108.

Regenwetter, M., J. Dana, and C. P. Davis-Stober (2010). Testing Transitivity of Preferences on Two-Alternative Forced Choice Data. *Frontiers in Psychology 1*(148), 1–15.

Regenwetter, M., J. Dana, and C. P. Davis-Stober (2011). Transitivity of Preferences. *Psychological Review 118*(1), 42–56.

Shafer, W. J. (1974). The Nontransitive Consumer. *Econometrica 42*, 913–919.

Sopher, B. and G. Gigliotti (1993). Intransitive Cycles: Rational Choice or Random Error? An Answer Based on Estimation of Error Rates with Experimental Data. *Theory and Decision 35*(3), 311–336.

Starmer, C. (1999). Cycling with Rules of Thumb: An Experimental Test for a New Form of Non- Transitive Behaviour. *Theory and Decision 46*, 141–158.

Starmer, C. (2000). Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice Under Risk. *Journal of Economic Literature 38*(2), 332–382.

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

Tversky, A. (1969). Intransitivity of Preferences. *Psychological Review 76*, 31–48.

Tversky, A. and D. Kahneman (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty 5*(4), 297–323.

Webb, R. (2019). The (Neural) Dynamics of Stochastic Choice. *Management Science 64*(1), 230–255.

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

Wilding, J. (1971). The Relation between Latency and Accuracy in the Identification of Visual Stimuli. I. The Effects of Task Difficulty. *Acta Psychologica 35*(5), 378–398.

APPENDICES

A Proof of Theorem 1

Proof. Let an SCF-RT (p, f) including data on a choice (x, y) be rationalized by an RCM-CF $(v, \tilde{v}, r)$. Let $G(x, y)$ denote the cumulative distribution function of $g(x, y)$, the density function of $\tilde{v}(x, y)$.

First we remark that

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

To see this, note that, by Definitions 1, 2, and 3, $p(y, x) = G(x, y)(0)$, $p(x, y) = 1 - G(x, y)(0)$, $F(x, y)(t) = (1 - G(x, y)(r^{-1}(t))/(1 - G(x, y)(0))$, and $F(y, x)(t) = G(x, y)(-r^{-1}(t))/G(x, y)(0)$. Thus,

$$\begin{aligned} p(y,x)F(y,x)(t) - p(x,y)F(x,y)(t) = G(x,y)(-r^{-1}(t)) - (1 - G(x,y)(r^{-1}(t)) = \\ G(x,y)(r^{-1}(t) + G(y,x)(-r^{-1}(t)) - 1. \end{aligned}$$

Second, by the integrated tail formula for expectations (Lo, 2019), and since $G(x, y)$ is the cumulative distribution function of the real-valued random variable $\tilde{v}(x, y)$,

$$\begin{aligned} v(x,y) = E[\tilde{v}(x,y)] = -\int_{-\infty}^{0} G(x,y)(v)dv + \int_{0}^{+\infty} (1 - G(x,y)(v))dv = \\ -\int_{0}^{+\infty} G(x,y)(-v)dv + \int_{0}^{+\infty} (1 - G(x,y)(v))dv = \\ \int_{0}^{+\infty} (1 - G(x,y)(v) - G(x,y)(-v))dv \end{aligned}$$

For any $v > 0$, let $t = r(v)$. By the remark above, the condition that $F(y, x)(t) \leq (p(y, x)/p(x, y))\, F(x, y)(t)$ can be rewritten as

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

for any v with $t = r(v) > 0$. This inequality then also holds for $v = 0$ by continuity. For any v with $r(v) = 0$, $G(x, y)(v) = 1$ and $G(x, y)(-v) = 0$, as otherwise the corresponding RCM-CF would generate an atom at response time zero. Hence $G(x, y)(v) + G(x, y)(-v) = 1$ in this case. It follows that the term in the final integral above is always positive, thus $v(x, y) \geq 0$ and the conclusion follows.

If, additionally, the inequality $F(y, x)(t) \leq (p(y, x)/p(x, y))\, F(x, y)(t)$ is strict for some t, it must be strict for a nonempty interval by continuity, implying $v(x, y) > 0$. □

B Empirical Implementation of the TWT Method

To reveal preferences using the TWT method, we proceed as follows (see also Alós-Ferrer et al., 2021). Given a (panel) dataset containing binary choices and the associated response times, we first compute the choice frequencies for each pair (x, y) and participant i, $p_i(x, y)$.

The second step is to estimate the density of the distribution of response times of i and (x, y). We used log-transformed response times for convenience, to avoid boundary problems and increase visibility of graphical illustrations, but this is largely inconsequential for the actual analysis. As in Alós-Ferrer et al. (2021), the kernel density estimates were performed in *Stata* using the *akdensity* function, which delivers CDFs as output, defined on a grid of time points t. This is a single-line command in Stata and other standard statistical packages (R: *Epa.kernel*, Mathlab: *ksdensity*). The estimates use an Epanechnikov kernel with optimally chosen non-adaptive bandwidth. If an option is chosen only once (and hence only one response time is available) an optimal bandwidth cannot be determined endogenously, so we set it to 0.1, yielding a distribution function close to a step function at the observed response time.

Once the CDF estimates $CDF_i(x,y)$ are obtained, we proceed to checking the condition of Theorem 1 for each participant i and choice pair (x,y). Specifically, we compute

$$H_i(x,y)(t) = p_i(x,y) \cdot CDF_i(x,y)(t) - p_i(y,x) \cdot CDF_i(y,x)(t)$$

for all t in the grid and check whether this expression is always negative or always positive (it is enough to check whether the minimum and maximum have the same sign). If $H_i(x,y)(t)$ is always positive (resp. always negative), a preference for x over y (resp. for y over x) is revealed for participant i. For each dataset, this can also be illustrated graphically by plotting the function H_i.

ONLINE APPENDIX FOR: Identifying Nontransitive Preferences

Carlos Alós-Ferrer*[1], Ernst Fehr[2], and Michele Garagnani[3]

[1]Lancaster University Management School
[2]Department of Economics, University of Zurich
[3]Department of Finance, University of Melbourne

May 13, 2024

A Previous Evidence on Nontransitivities

Systematic empirical evidence on transitivity violations goes back to May (1954), who collected choice data for pairs of hypothetical marriage partners described according to intelligence, looks, and wealth. However, the evidence was in the form of nontransitive cycles when the choices of all participants were aggregated, and hence reduces to the well-known observation that Condorcet cycles might appear when transitive preferences are aggregated. Actual evidence on nontransitive preferences at the individual level was first presented by Tversky (1969), using binary choices among simple monetary lotteries and also among hypothetical job applicants. Almost all participants displayed at least one weak stochastic transitivity violation. These descriptive findings were subsequently replicated (Montgomery, 1977; Lindman and Lyons, 1978; Budescu and Weiss, 1987), but the later literature cast doubts on the strength of the evidence. Iverson and Falmagne (1985) reanalyzed the data of Tversky (1969) and argued that the evidence was compatible with transitive preferences and noisy choices. They further criticized the original work's statistical analysis and found that only one of Tversky's participants significantly violated transitivity using likelihood ratio tests, which of course implicitly assume (a particular shape of) noise in actual choices. It has also been criticized that participants in Tversky (1969) were pre-selected.

Later empirical demonstrations of nontransitive choice have been similarly criticized, the core argument frequently being that data might be compatible with transitive but noisy behavior. For example, Loomes et al. (1989, 1991) argued that the classical preference reversal phenomenon (Lichtenstein and Slovic, 1971; Grether and Plott, 1979;

*Corresponding author. Lancaster University Management School, LA1 4YX Lancaster (U.K.). E.mail: c.alosferrer@lancaster.ac.uk

Tversky and Thaler, 1990), where choices systematically contradict elicited (monetary) valuations, might be due to transitivity violations. That is, actual nontransitive choices might build a preference cycle where a lottery A is preferred to a lottery B and this second lottery is (of course) revealed indifferent to its own certainty equivalent, but the latter is strictly preferred to the certainty equivalent of A. However, Sopher and Gigliotti (1993), in a replication of Loomes et al. (1991), estimated an econometric model of choice with a specific structure of random errors, and could not reject the null hypothesis of transitive preferences and noisy choices. On the other hand, Starmer and Sugden (1998) further replicated the work in Loomes et al. (1991) and observed the same cycling asymmetries, suggesting that those are unlikely to be due to noise.

Regenwetter et al. (2010, 2011) argued that violations of transitivity are better analyzed through violations of the triangle inequality, $p(x, y) + p(y, z) - p(x, z) \leq 1$ (Marschak, 1960; Block and Marschak, 1960), instead of violations of Weak Stochastic Transitivity. Those works found that the triangle inequality is often satisfied in (many) existing publications, even when WST is violated. Cavagnaro and Davis-Stober (2014) argued that the the behavior of the tested populations can be best described by a mixture of different models of choice, with the resulting estimates suggesting that the majority (but not all) of the people might satisfy transitivity.

Recent studies, however, keep bringing up empirical evidence which might indicate violations of transitivity. Butler and Pogrebna (2018) provided new empirical evidence using both WST and the triangle inequality. Their evidence showed that cycles can be the modal preference patterns over simple lotteries even after considering transitive, stochastic models. Their choices were designed to reproduce the "paradox of nontransitive dice," where a heuristic which favors the option (within a pair) with the largest probability to beat the alternative produce cyclical choices (Savage Jr., 1994). As in previous cases, however, critical work was close on the heels of Butler and Pogrebna (2018). Specifically, Birnbaum (2023) argued that tests of Weak Stochastic Transitivity and the triangle inequality do not provide a method to compare transitive and nontransitive models that allow mixtures of preference patterns and random errors. Birnbaum (2020) re-analyzed the data of Butler and Pogrebna (2018) using a "true and error" model (Remark 1 in the main text) and still found evidence for significant transitivity violations, but the latter are incompatible with the explanation proposed by Butler and Pogrebna (2018) (see, however, Butler, 2020).

Following up on the general argument that apparent violations of transitivity could be rationalized by transitive models if one focuses on choice frequencies only, Fosgerau and Rehbeck (2023) study a specific class of models of nontransitive choice where choice probabilities reflect strength of preference and characterize under which conditions they can be rationalized within a class of noisy choice models with transitive preferences.

Observed violations of transitivity, whatever their origin, seem to be relatively stable. For example, Davis-Stober et al. (2019) and Park et al. (2019) report that neither age nor, surprisingly, alcohol intoxication seem to play a major role in transitivity violations for decisions under risk. Non-transitive choices have also been observed in other domains. Li and Loomes (2022) report a substantial level of nontransitive choices in respondents' intertemporal decisions, i.e. decisions between pairs of monetary amounts to be received at different points in time (see also Tversky et al., 1990). Birnbaum and Schmidt (2008) find some evidence for transitivity violations for choices under uncertainty, albeit for a limited number of participants. Moreover, people frequently violate transitivity when choosing between multi-attribute consumers' products (sound systems, flight plans, and software packages; e.g. Lee et al., 2009; Müller-Trede et al., 2015; Lee et al., 2015). Naturally, there are also some domains where evidence is less robust, e.g. for hypothetical alternative treatments in the health domain (Schmidt and Stolpe, 2011), or when choosing between potential sexual partners (Hatz et al., 2020). Finally, violations of transitivity are no exception to the rule that few behaviors, if at all, are uniquely human: honey bees and gray jays have been shown to violate transitivity when foraging for food (Shafir, 1994; Waite, 2001), and Túngara frogs behave nontransitively when making mating choices (Natenzon, 2019).

A few contributions have also tested for particular forms of transitivity violations. For instance, Starmer and Sugden (1998) documented transitivity violations which might contradict a number of explanations, including regret theory. Starmer (1999) tested for transitivity violations which might be compatible with the "editing phase" of original prospect theory (Kahneman and Tversky, 1979). We refer the reader to Starmer (2000) for a discussion.

We remark that, in this work, we follow the literature which favors testing transitivity violations using binary choice probabilities instead of choice patterns (e.g., Birnbaum, 2020). For a discussion of these two alternative approaches, we refer the reader to Cavagnaro and Davis-Stober (2014) and Butler (2020). This is a natural choice given our theoretical framework, which reveals preferences using binary choices. Moreover, the two approaches have been shown to provide largely consistent evidence (e.g., Butler and Pogrebna, 2018; Birnbaum, 2020).

Part of the previous literature has concentrated on fitting data to particular models and comparing the fit of transitive and nontransitive models in "horse race" exercises. This approach is incomparable to ours, since we identify choice patterns that cannot be explained by *any* model of transitive preferences with behavioral noise, in the sense of Section 2 in the main text. However, the overall message of our findings, namely that there are persistent transitivity violations but a majority of choice combinations respect transitivity, is compatible with the recent literature, which finds consistent support for nontransitive models of choice.

For example, using true and error models (recall Remark 1 in the main text), Birnbaum (2023) reports that most participants in the experiment of Butler and Pogrebna (2018) made decisions consistent with transitivity, but 7 out of 22 (about 30%) showed evidence of nontransitive preference patterns at least part of the time. Brown et al. (2015), reanalyzing the data from KTHDP, find that 7 out of 30 participants were best described by models which allow for intransitivities, while 8 participants were best explained by a trembling hand model (again, recall Remark 1 in the main text) and 6 other participants were best explained by a stochastic preference model (hence equivalent to a classical, additive RUM). Ranyard et al. (2020), reanalyzing the same dataset, found that a model accounting for violations of WST (based on the additive difference model of Tversky, 1969) was a good fit for 14 of the 30 participants.

Needless to say, this section is not and cannot be a complete review of the literature on transitivity violations. We refer the reader to the recent review of Ranyard et al. (2020), who also estimated a simplified additive-difference model based on the processing of alternative dimensions (following Tversky, 1969). Similarly to Regenwetter et al. (2010, 2011), Ranyard et al. (2020) argue that people seem to behave according to different models of choice, and many individuals are best explained by models which do violate transitivity.

B Previous Deterministic, Nontransitive Models

For multidimensional alternatives, $x = (x_1, \ldots, x_n)$, Tversky (1969) introduced the *additive difference model* with the explicit purpose of studying nontransitivities. This model postulates that $x \succeq y$ if and only if

$$\sum_{i=1}^{n} \phi_i(u_i(x_i) - u_i(y_i)) \geq 0$$

where u_i are real-valued factor utilities, and ϕ_i are skew-symmetric ($\phi_i(-r) = -\phi_i(r)$), increasing and continuous real-valued functions. This expression becomes an example of a function v as in Shafer (1974) for the multidimensional case.

In the domain of decisions under risk (lottery choice), Kreweras (1961) and Fishburn (1982, 1984, 1986), among others, studied skew-symmetric bilinear (SSB) representations, which reduce choice to the sign of a function v as in Shafer (1974) with the additional the requirement that v is linear in both arguments. Specifically, let L_1, L_2 be simple lotteries on the set of outcomes X, i.e. $L_1(x), L_2(x)$ denote the respective probabilities of outcome x and those are only positive for finitely many outcomes. A function v defined on outcomes can be extended bilinearly to simple lotteries by

$$V^{SSB}(L_1, L_2) = \sum_{x \in X} \sum_{y \in X} L_1(x) L_2(y) \cdot v(x, y).$$

so that L_1 is weakly preferred to L_2 if and only $V^{SSB}(L_1, L_2) \geq 0$. This generalizes expected utility, since if $v(x, y) = u(x) - u(y)$ for a utility function u on X, then $V^{SSB}(L_1, L_2) = \sum_{x \in X} L_1(x) u(x) - \sum_{y \in X} L_2(y) u(y)$. However, the SSB form does not require transitivity and indeed allows for preference cycles and violations of the independence axiom (see Fishburn, 1988 for an axiomatization of SSB nontransitive preferences).

That is, the function V^{SSB} is a particular example of the approach of Shafer (1974) for a space of lotteries.

Some other prominent theories have incorporated regret and salience in decision making under risk by capturing these phenomena in a skew-symmetric function over outcomes and then extending that function to lotteries in a manner similar to SSB models. These models, however, are formulated in terms of *acts* (Savage, 1954), that is, mappings from a set of states to outcomes, and hence it is better to change notation at this point. Let the (finite) set of states be denoted by S, and let $p(s)$ denote the probability of a state $s \in S$. A lottery L^x is then a vector of outcomes $(x_s)_{s\in S}$, with the interpretation that outcome x_s obtains if state s occurs. Loomes and Sugden (1982) introduced *regret theory* as a particular model allowing for transitivity violations in the risk domain (see Starmer, 2000, for a summary). Diecidue and Somasundaram (2017) showed that regret theory deviates from expected utility only by relaxing transitivity. Loomes and Sugden (1987) later extended this framework to *generalized regret theory.* This theory considers monetary consequences, $X \subseteq \mathbb{R}$, and starts out by postulating a real-valued, two-argument function M, so that if $x, y \in X$, $M(x, y)$ is interpreted as the utility of choosing x net of the regret associated with missing out on y. Then $M(x, y)$ becomes the basis for defining the function v^R by $v^R(x, y) = M(x, y) - M(y, x)$ which is immediately skew-symmetric and hence a particular case of the approach of Shafer (1974) for the space of outcomes. Analogously to SSB models, but within the formalization of lotteries as acts, a lottery L^x is weakly preferred to a lottery L^y if and only if $V^R(L^x, L^y) \geq 0$, where

$$V^R(L^x, L^y) = \sum_{s\in S} p(s)v^R(x_s, y_s).$$

Loomes and Sugden (1987) further impose several assumptions on v^R, namely that $v^R(x, y) \geq 0$ if and only if $x \geq y$ (so that, for outcomes, more is better), that $v^R(x, z) > v^R(y, z)$ (resp. $<, =$) if and only if $x > y$ (resp. $<, =$), and a "regret aversion" assumption stating that $v^R(x, z) > v^R(x, y) + v^R(y, z)$ whenever $x > y > z$, meaning that large post-decision regrets are worse than the sum of step-wise, smaller regrets. In particular, skew symmetry and these conditions imply that $v(x, y) > 0$ if $x > y$, $v(x, y) < 0$ if $x < y$, and $v(x, x) = 0$, for any outcomes x, y.

The comparison of regret theory and SSB theory is obscured by the fact that the former is formulated in terms of lotteries as acts, while the latter is formulated in terms of lotteries as probability distributions. Loomes and Sugden (1987) show that, for stochastically independent lotteries (where the set of states can be seen as a product of lottery-specific sets of states), generalized regret theory is equivalent to SSB theory. Again, the function V^R becomes a particular example of the approach of Shafer (1974) for a space of lotteries.

Bordalo et al. (2012, 2013) introduced *salience theory* by postulating a *symmetric* function σ, i.e. $\sigma(x, y) = \sigma(y, x)$ for all $x, y \in X \subseteq \mathbb{R}$, with the interpretation that for a lottery pair (L^x, L^y), $\sigma(x_s, y_s)$ is the salience of the state s. This function is assumed to fulfill a number of properties capturing the idea of salience. In a "smooth" version of the theory, salience values are transformed through an increasing, real-valued function f which preserves salience rankings as derived from σ, yielding[1]

$$q_s(L^x, L^y) = \frac{f(\sigma(x_s, y_s))}{\sum_{r \in S} f(\sigma(x_r, y_r))}.$$

The decision maker then attaches a value to lottery L^x which depends on the alternative lottery L^y,

$$U^{ST}(L^x|L^y) = \sum_{s \in S} q_s\left(L^x, L^y\right) u(x_s)$$

where u is strictly increasing with $u(0) = 0$.

Although (smooth) salience theory appears functionally different from generalized regret theory and SSB models, it is worth observing that there is a relation. Under salience theory, a lottery L^x is weakly preferred to a lottery L^y if and only if $V^{ST}(L^x, L^y) \geq 0$, where

$$V^{ST}(L^x, L^y) = \sum_{s \in S} p(s) f(\sigma(x_s, y_s)) \left[u(x_s) - u(y_s)\right].$$

This already shows that regret theory is a further particular case of the approach of Shafer (1974) for a space of lotteries. Herweg and Müller (2021) further observe that the two-argument function on outcomes w^{ST} defined by $w^{ST}(x, y) = f(\sigma(x, y)) \left[u(x) - u(y)\right]$ is skew symmetric, and hence salience theory can be written in the same terms as generalized regret theory. Herweg and Müller (2021) also show that, assuming continuity of u and f, the assumptions of (smooth) salience theory imply those of generalized regret theory, that is, one can view salience theory as a particular case of the latter, and hence (for stochastically independent lotteries) as a particular case of SSB theory. Interestingly, the original regret theory of Loomes and Sugden (1982), which was a more specific model, turns out to be a particular case of salience theory if an additional, mild condition is imposed (Herweg and Müller, 2021, Theorem 2).[2]

All theories discussed above obviously allow for nontransitivities in lottery choice, since they can be described as special cases of the fundamental representation of Shafer (1974).[3] That is, ultimately they provide a (structural, parametric) functional form for

[1] Bordalo et al. (2012) also provide a *rank-based* version of salience theory with similar insights. This version is analytically more tractable for specific applications, but creates discontinuities in valuations (Kontek, 2016).

[2] It can be shown that generalized regret theory (and hence smooth salience theory) fulfill a weaker version of transitivity, called *dominance transitivity* by Diecidue and Somasundaram (2017): if L^x strictly dominates L^y (yields better outcomes for all states, and strictly better for at least some states) and the latter is preferred to L^z, then L^x must be strictly preferred to L^z (and analogously if L^x is preferred to L^y and the latter strictly dominates L^z). This rather weak condition seems to be the only systematic constraint on the kind of transitivity violations that these models can generate.

[3] Other models that allow for transitivity violations include lexicographic semiorders (Hausner, 1954; Fishburn, 1971), similarity theory (Fishburn, 1991; Leland, 1994, 1998), the context-dependent model of the gambling effect (Bleichrodt and Schmidt, 2002), and the stochastic difference model of González-Vallejo (2002).

a function $V(\cdot,\cdot)$ defined on a specific space, while the general approach of Shafer (1974) allows for any skew-symmetric function.

C Robustness Analysis: No Effects of Attention

To check whether possible distraction as the experiments progressed influenced the results, we split each dataset into "early" and "late" decisions. In KTHDP we simply split the data of each individual in two halves. For DSBC, since data were collected in four sessions, we compare the data collected in the first half of the sessions to the data from the second half of the sessions, i.e. the split is carried out for each session. There are no statistically significant differences in neither DSBC nor KTHDP when comparing the first and the second half of the sessions. Specifically, this holds for the following comparisons.

- Individual proportion of Revealed Transitivity Violations (RTVs) (KTHDP: first half 12.24% vs. second half 14.63%, WSR test, $N = 30$, $z = -0.575$, $p = 0.5745$; DSBC: 19.87% vs. 18.84%, WSR test, $N = 60$, $z = 1.207$, $p = 0.2306$).

- Individual proportion of pairs where preferences are revealed according to Theorem 1 (KTHDP: first half 77.85% vs. second half 75.99%, WSR test, $N = 30$, $z = 1.602$, $p = 0.1112$; DSBC first half 56.52% vs. second half 56.82%, WSR test, $N = 60$, $z = -0.961$, $p = 0.3411$).

- Individual proportion of WST violations (KTHDP: first half 14.86% vs. second half 16.00%, WSR test, $N = 30$, $z = -0.460$, $p = 0.6581$; DSBC: 21.10% vs. 20.41%, WSR test, $N = 60$, $z = 1.318$, $p = 0.1903$).

Additionally, KTHDP included choice pairs were one alternative first-order stochastically dominated the other. Violations of stochastic dominance could be an indicator of individual-level inattention or distraction. However, we found no significant correlation between the individual proportion of dominated choices and the three main outcomes given above, that is,

- the individual proportion of WST violations (Spearman's $N = 30$, $\rho = 0.2806$, $p = 0.1332$),

- the individual proportion of Revealed Transitivity Violations (RTVs) ($N = 30$, $\rho = 0.2140$, $p = 0.2562$),

- or the individual proportion of pairs where preferences are revealed according to Theorem 1 (Spearman's $N = 30, \rho = -0.2810, p = 0.1325$).

Also, the individual proportions of dominated choices in this dataset are not statistically significantly different between the first and the second half of the session (first half 22.27% vs. second half 22.06%, WSR test, $N = 29$, $z = 0.303$, $p = 0.7698$).

D Robustness Analysis: Time Limits and Lottery Formats

In DSBC two *within-subject* treatments were implemented, time limits vs. no time limits and pie vs. bar lottery format. We can hence investigate the possible influence of these manipulations on our results.

There were no significant differences between sessions with and without time limits for

- the individual proportion of WST violations (time limit 20.32% vs. no time limit 21.21%; $WRSN = 60, z = -0.578, p = 0.5671$),
- the individual proportion of Revealed Transitivity Violations (RTVs) (time limit 19.03% vs. no time limit 19.15%; WRS $N = 60$, $z = -0.129$, $p = 0.9011$),
- or the individual proportion of pairs where preferences are revealed according to Theorem 1 (time limit 56.13% vs. no time limit 57.16%; WRS, $N = 60$, $z = -0.942$, $p = 0.3505$).

We observe that using the bar representation is associated with a higher proportion of revealed preferences (59.87%) compared to the pie representation (53.84%; WRS $N = 60$, $z = 3.872$, $p < 0.001$). Pie representations also do lead to a larger proportion of WST violations compared the bar representations, although the comparison misses significance at the 5% level (21.35% vs. 20.19%; WRS $N = 60$, $z = 1.716$, $p = 0.0866$). There are, however, no significant differences in the proportion of RTVs when comparing pie and bar representations (18.52% vs. 20.04%; WRS $N = 60$, $z = -0.648$, $p = 0.5222$).

E Characteristics of Nontransitive Cycles

Our analysis in the main text (Section 3) shows the existence of transitivity violations which are not due to noise. A natural question is whether specific collections of lottery choices give rise to such violations often. To answer this question, we reanalyze the data taking individual cycles of alternatives as the unit of observation. That is, in each dataset and for each cycle of alternatives, we compute the percentage of participants who display either an RTV or a WST violation.

The left-hand panel of Figure 1 represents the distribution of the proportion of participants displaying RTVs across cycles of alternatives, computed over all participants for which the cycle was revealed (DSBC: mean 18.93%, median 17.65%, SD=11.69, min 0.00%, max 60.00%; KTHDP: mean 11.22%, median 0.00%, SD=20.55, min 0.00%, max 100.00%). The right-hand panel shows the distribution of the proportion of participants displaying WST violations across cycles of alternatives, computed over all participants (DSBC: mean 20.56%, median 21.67%, SD=6.51, min 0.00%, max 36.67%; KTHDP: mean 15.42%, median 0.00%, SD=20.03, min 0.00%, max 80.00%).[4] As can be seen in

[4]Note that for DSBC the average is computed over $N = 60 \times 4$ observations, as each participant made the same choices in four different conditions.

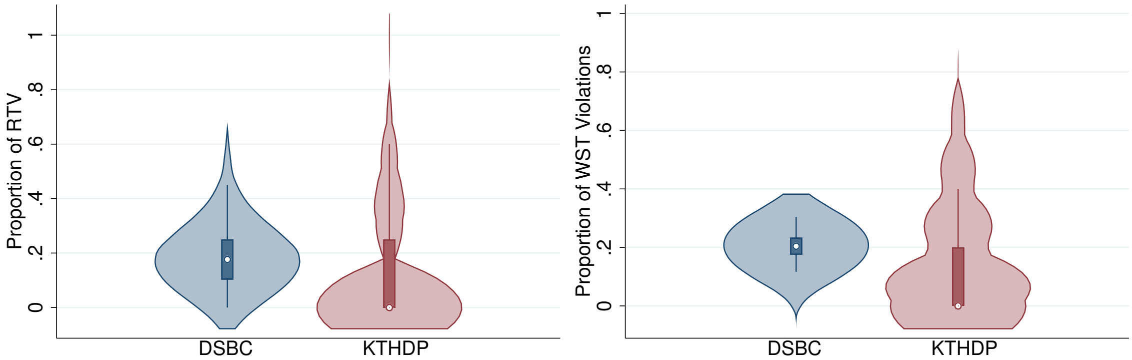

Figure 1: Distribution of the proportion of subjects displaying RTVs (on the left; relative to the set of subjects for which preferences are revealed in the corresponding cycle of alternatives) and WST violations (on the right; relative to all subjects) per each cycle of alternatives.

the figure, the support of the distributions range from zero to relatively large numbers. That is, some cycles of alternatives involve next to no violations while others involve nontransitive choices for a sizeable part of the experiment's participants.

To single out which constellations of choices produce a particularly large proportion of violations, we then look at the cycles of alternatives which entail the most transitivity violations. Table 1 lists the ten cycles (for both datasets) with the largest proportion of RTVs, computed as the percentage of people for which the cycle of alternatives was revealed who displayed an RTV. For DSBC, those range from 48% to 58%, and all of them correspond to WST violations for at least a quarter of the sample. Notably, all ten cycles involve just the five following lotteries (out of the nine in the experiment), which correspond to the left-hand subset in Figure 1(left) in the main text.

$$x_1 = \left(\$25.43, \frac{7}{24}\right), \; x_2 = \left(\$24.16, \frac{8}{24}\right), \; x_* = \left(\$22.89, \frac{9}{24}\right),$$
$$x_3 = \left(\$21.62, \frac{10}{24}\right), x_4 = \left(\$20.35, \frac{11}{24}\right)$$

The fact that the most common transitivity violations in DSBC all involve the left-hand subset in Figure 1(left) in the main text, and none of them involves the lotteries in the right-hand set, is particularly revealing. The differences in outcomes across similar lotteries in the right-hand set are noticeably larger (between \$3.13 and \$4.96) than those for the other set (all \$1.27), while differences in probabilities are always 1/24 in both sets. That is, the most frequent nontransitivities involve choices whose evaluations are presumably closer, i.e. such that the strength of preference is smaller. If one used WST or a similar measure as a criterion for detecting nontransitivites, standard psychometric effects (error rates are larger for closer valuations) would suggest that the increase in nontransitivities is merely due to increased noise. However, our approach through RTVs

Table 1: The ten cycles in DSBC and KTHDP with the most transitivity violations. The second column indicates the proportion of experimental participants displaying the RTV given in the first column, computed over all participants for which the corresponding cycle of altenratives was revealed (numbers in brackets indicate how the proportion is computed). The third column indicates the proportion of participants (out of 4×60 for DSBC, out of 30 for KTHDP) displaying a WST violation for the cycle of alternatives.

Cycle	People with RTV	People with WST
DSBC		
$x_* \succ x_4 \succ x_3 \succ x_*$	58.33% (28/48)	30.00% (72)
$x_* \succ x_4 \succ x_3 \succ x_2 \succ x_*$	54.55% (24/44)	35.00% (84)
$x_* \succ x_1 \succ x_4 \succ x_3 \succ x_*$	50.00% (12/24)	25.00% (60)
$x_* \succ x_4 \succ x_2 \succ x_3 \succ x_*$	53.85% (28/52)	31.67% (76)
$x_* \succ x_2 \succ x_3 \succ x_4 \succ x_*$	57.14% (32/56)	33.33% (80)
$x_* \succ x_2 \succ x_3 \succ x_1 \succ x_4 \succ x_*$	47.62% (40/84)	36.67% (88)
$x_* \succ x_4 \succ x_1 \succ x_3 \succ x_2 \succ x_*$	50.00% (24/48)	35.00% (84)
$x_* \succ x_1 \succ x_4 \succ x_2 \succ x_3 \succ x_*$	50.00% (12/24)	26.67% (64)
$x_* \succ x_4 \succ x_1 \succ x_2 \succ x_3 \succ x_*$	53.85% (28/52)	35.00% (84)
$x_* \succ x_4 \succ x_2 \succ x_1 \succ x_3 \succ x_*$	57.14% (32/56)	33.33% (80)
KTHDP		
$y_2 \succ y_4 \succ y_5 \succ y_2$	66.67% (12/18)	40.00% (12)
$y_2 \succ y_3 \succ y_5 \succ y_4 \succ y_2$	100.00% (6/6)	60.00% (18)
$y_3 \succ y_1 \succ y_2 \succ y_4 \succ y_3$	66.67% (12/18)	60.00% (18)
$y_3 \succ y_4 \succ y_5 \succ y_1 \succ y_3$	66.67% (12/18)	40.00% (12)
$y_4 \succ y_1 \succ y_2 \succ y_3 \succ y_4$	66.67% (12/18)	60.00% (18)
$y_1 \succ y_4 \succ y_3 \succ y_2 \succ y_5 \succ y_1$	75.00% (9/12)	30.00% (9)
$y_3 \succ y_1 \succ y_2 \succ y_4 \succ y_5 \succ y_3$	66.67% (12/18)	60.00% (18)
$y_3 \succ y_4 \succ y_5 \succ y_1 \succ y_2 \succ y_3$	66.67% (12/18)	40.00% (12)
$y_4 \succ y_1 \succ y_2 \succ y_3 \succ y_5 \succ y_4$	66.67% (12/18)	60.00% (18)
$y_4 \succ y_5 \succ y_2 \succ y_1 \succ y_3 \succ y_4$	66.67% (12/18)	40.00% (12)

has disentangled preferences from noise. Thus, the data suggests that the increase in nontransitivities is due to the fact that evaluations are close, but not because this results in noisier choices. Rather, it appears that empirical transitivity violations are more frequent when they result from a gradual chain of small changes in the options. Specifically, many of the examples in Table 1 suggest that small tradeoffs, which are possible when lottery attributes are close enough, do not scale up monotonically. For example, consider the shortest cycle for DSBC in Table 1, which is also the one with the largest proportion of RTV violations, $x_* \succ x_4 \succ x_3 \succ x_*$. Twice along this cycle ($x_4 \succ x_3 \succ x_*$), the decision maker accepts a one-step decrease in monetary payoff (\$1.27) in exchange for a one-step increase in probability (1/24). Then, however, the same decision maker accepts a two-steps decrease in probability (2/24) in exchange for a two-step increase in monetary payoff (\$2.54). The exact same phenomenon appears in the cycles $x_* \succ x_4 \succ x_3 \succ x_2 \succ x_*$, $x_* \succ x_2 \succ x_3 \succ x_4 \succ x_*$, and (rewritten) $x_1 \succ x_4 \succ x_3 \succ x_* \succ x_1$, with three one-step tradeoffs being reversed by a three-step tradeoff in the opposite direction, and similar but more complex patterns can be seen in

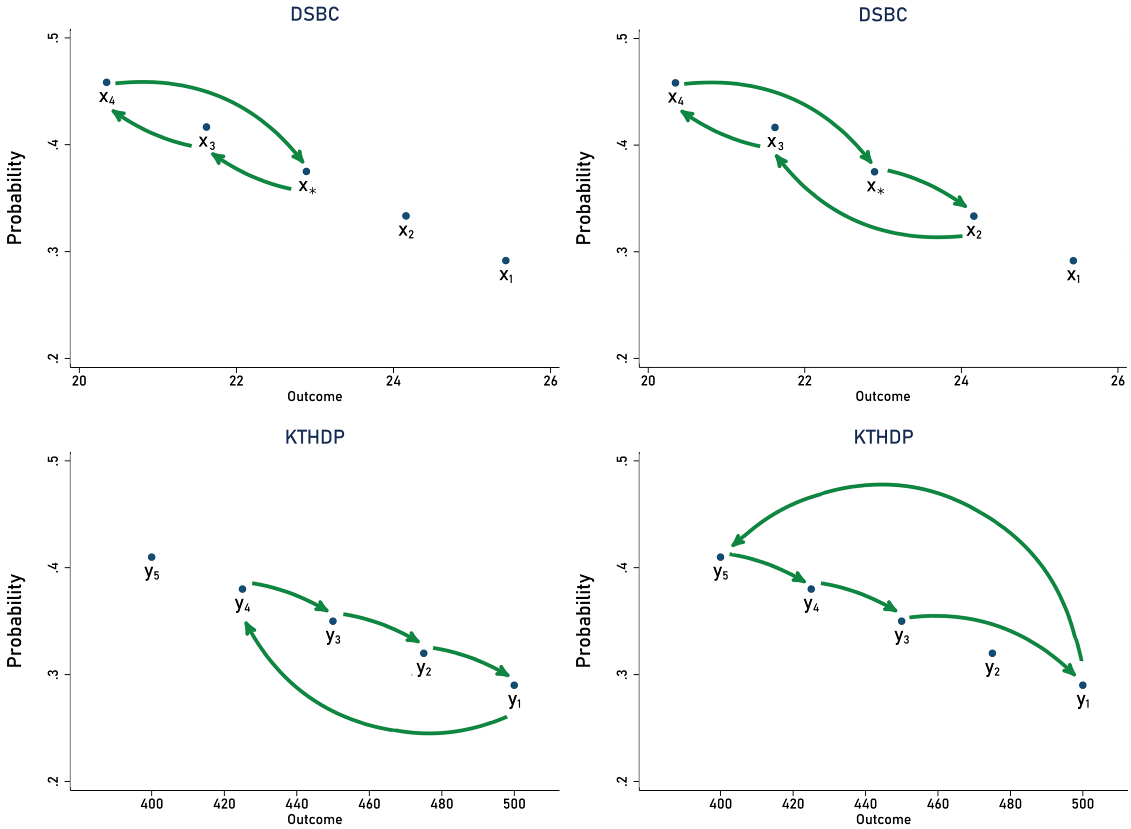

Figure 2: Graphical representation of some of the most common preference cycles in the datasets. All lotteries have a single non-zero outcome, depicted in the (outcome, probability) space. Arrows indicate preference, i.e. $x \to y$ means $y \succ x$. The two upper pictures are from the DSBC data, the two lower ones from KTHDP.

the longer cycles. The two top panels of Figure 2 give a graphical representation of two of these examples.

For KTHDP, the proportion of RTVs among revealed cycles of alternatives for the ten topmost ones is always above two thirds. corresponding to between 40% and 60% WST violations in the overall sample. The cycles involve all five lotteries in KTHDP,

$$y_1 = (\$500, 0.29)\,,\ y_2 = (\$475, 0.32)\,,\ y_3 = (\$450, 0.35)\,,$$
$$y_4 = (\$425, 0.38)\,,\ y_5 = (\$400, 0.41)$$

The same phenomenon is observed in several of the KTHDP cycles. For example, in the cycle $y_4 \succ y_1 \succ y_2 \succ y_3 \succ y_4$, three times in a row the decision maker accepts a one-step reduction in probability (0.03) in exchange for a one-step increase in monetary payoff ($25), but then undoes it by accepting a three-step reduction in monetary payoff ($75) in exchange for a three-step increase in probability (0.09). A similar pattern can be seen in the cycle $y_3 \succ y_4 \succ y_5 \succ y_1 \succ y_3$, and similar phenomena appear in several of the longer cycles. The two bottom panels of Figure 2 give a graphical representation of two of these examples.

We remark that the nontransitivities we discuss above (chains of small changes which do not scale up) are not related to the classical idea of "nontransitive indifference." The latter argument points out that apparent nontransitivities might be frequently observed whenever decision makers are close to indifference, because (by the psychometric effect) noise is maximal in this case. However, our approach disentangles preferences from noise. If a decision maker displayed an indifference cycle, many apparent nontransitivities would be observed in the data, but, since the underlying preferences are not strict, if follows that Theorem 1 cannot reveal a strict preference; in practice, the sufficient condition would fail for at least one of the involved pairs. Hence, those apparent transitivity violations would not be Revealed Transitivity Violations. In other words, our approach disentangles genuine transitivity violations from "noisy ones" as would be the case for those arising from nontransitive indifference.

References

Birnbaum, M. H. (2020). Reanalysis of Butler and Pogrebna (2018) Using True and Error Model. *Judgment and Decision Making 15*(6), 1044–1051.

Birnbaum, M. H. (2023). Testing Transitivity of Preference in Individuals. *Decision 10*(2), 153–180.

Birnbaum, M. H. and U. Schmidt (2008). An Experimental Investigation of Violations of Transitivity in Choice Under Uncertainty. *Journal of Risk and Uncertainty 37*(1), 77–91.

Bleichrodt, H. and U. Schmidt (2002). A Context-Dependent Model of the Gambling Effect. *Management Science 48*(6), 802–812.

Block, H. D. and J. Marschak (1960). Random Orderings and Stochastic Theories of Responses. In I. Olkin (Ed.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, pp. 97–132. Stanford: Stanford University Press.

Bordalo, P., N. Gennaioli, and A. Shleifer (2012). Salience Theory of Choice under Risk. *Quarterly Journal of Economics 127*(3), 1243–1285.

Bordalo, P., N. Gennaioli, and A. Shleifer (2013). Salience and Consumer Choice. *Journal of Political Economy 121*(5), 803–843.

Brown, N., C. P. Davis-Stober, and M. Regenwetter (2015). Commentary: "Neural Signatures of Intransitive Preferences". *Frontiers in Human Neuroscience 9*, 509.

Budescu, D. V. and W. Weiss (1987). Reflection of Transitive and Intransitive Preferences: A Test of Prospect Theory. *Organizational Behavior and Human Decision Processes 39*(2), 184–202.

Butler, D. (2020). Intransitive Preferences or Choice Errors? A Reply to Birnbaum. *Judgment and Decision Making 15*(6), 1052–1053.

Butler, D. and G. Pogrebna (2018). Predictably Intransitive Preferences. *Judgment and Decision Making 13*(3), 217–236.

Cavagnaro, D. R. and C. P. Davis-Stober (2014). Transitive in Our Preferences, but Transitive in Different Ways: An Analysis of Choice Variability. *Decision 1*(2), 102–122.

Davis-Stober, C. P., D. M. McCarthy, D. R. Cavagnaro, M. Price, N. Brown, and S. Park (2019). Is Cognitive Impairment Related to Violations of Rationality? A Laboratory Alcohol Intoxication Study Testing Transitivity of Preference. *Decision 6*(2), 134–144.

Diecidue, E. and J. Somasundaram (2017). Regret Theory: A New Foundation. *Journal of Economic Theory 172*, 88–119.

Fishburn, P. C. (1971). A Study of Lexicographic Expected Utility. *Management Science 17*(11), 672–678.

Fishburn, P. C. (1982). Nontransitive Measurable Utility. *Journal of Mathematical Psychology 26*(1), 31–67.

Fishburn, P. C. (1984). SSB utility theory: An economic perspective. *Mathematical Social Sciences 8*(1), 63–94.

Fishburn, P. C. (1986). Ordered Preference Differences Without Ordered Preferences. *Synthese 67*(2), 361–368.

Fishburn, P. C. (1988). *Nonlinear Preference and Utility Theory.* Baltimore, Maryland: Johns Hopkins University Press.

Fishburn, P. C. (1991). Nontransitive Preferences in Decision Theory. *Journal of Risk and Uncertainty 4*(2), 113–134.

Fosgerau, M. and J. Rehbeck (2023). Nontransitive Preferences and Stochastic Rationalizability: A Behavioral Equivalence. Working Paper, arXiv:2304.14631.

González-Vallejo, C. (2002). Making Trade-Offs: A Probabilistic and Context-Sensitive Model of Choice Behavior. *Psychological Review 109*(1), 137–155.

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

Hatz, L. E., S. Park, K. N. McCarty, D. M. McCarthy, and C. P. Davis-Stober (2020). Young Adults Make Rational Sexual Decisions. *Psychological Science 31*(8), 944–956.

Hausner, M. (1954). Multidimensional Utilities. *Decision Processes*, 167–180.

Herweg, F. and D. Müller (2021). A Comparison of Regret Theory and Salience Theory for Decisions Under Risk. *Journal of Economic Theory 193*, 105226.

Iverson, G. and J.-C. Falmagne (1985). Statistical Issues in Measurement. *Mathematical Social Sciences 10*(2), 131–153.

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

Kontek, K. (2016). A Critical Note on Salience Theory of Choice Under Risk. *Economics Letters 149*, 168–171.

Kreweras, G. (1961). Sur une possibilité de rationaliser les intransitivités. In *La Décision, Colloques Internationaux du Centre National de la Recherche Scientifique*, pp. 27–32. Paris: Editions du Centre National de la Recherche Scientifique.

Lee, L., O. Amir, and D. Ariely (2009). In Search of Homo Economicus: Cognitive Noise and the Role of Emotion in Preference Consistency. *Journal of Consumer Research 36*(2), 173–187.

Lee, L., M. P. Lee, M. Bertini, G. Zauberman, and D. Ariely (2015). Money, Time, and the Stability of Consumer Preferences. *Journal of Marketing Research 52*(2), 184–199.

Leland, J. W. (1994). Generalized Similarity Judgments: An Alternative Explanation for Choice Anomalies. *Journal of Risk and Uncertainty 9*(2), 151–172.

Leland, J. W. (1998). Similarity Judgments in Choice Under Uncertainty: A Reinterpretation of the Predictions of Regret Theory. *Management Science 44*(5), 659–672.

Li, Z. and G. Loomes (2022). Revisiting the Diagnosis of Intertemporal Preference Reversals. *Journal of Risk and Uncertainty 64*(1), 1–23.

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

Lindman, H. R. and J. Lyons (1978). Stimulus Complexity and Choice Inconsistency Among Gambles. *Organizational Behavior and Human Performance 21*(2), 146–159.

Loomes, G., C. Starmer, and R. Sugden (1989). Preference Reversal: Information-Processing Effect or Rational Non-Transitive Choice? *Economic Journal 99*(395), 140–151.

Loomes, G., C. Starmer, and R. Sugden (1991). Observing Violations of Transitivity by Experimental Methods. *Econometrica 59*(2), 425–439.

Loomes, G. and R. Sugden (1982). Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty. *Economic Journal 92*(368), 805–824.

Loomes, G. and R. Sugden (1987). Some Implications of a More General Form of Regret Theory. *Journal of Economic Theory 41*(2), 270–287.

Marschak, J. (1960). Binary Choice Constraints on Random Utility Indicators. In K. J. Arrow (Ed.), *Stanford Symposium on Mathematical Methods in the Social Sciences*, pp. 312–329. Stanford, CA: Stanford University Press.

May, K. O. (1954). Intransitivity, Utility, and the Aggregation of Preference Patterns. *Econometrica 22*(1), 1–13.

Montgomery, H. (1977). A Study of Intransitive Preferences Using a Think Aloud Procedure. In *Decision Making and Change in Human Affairs*, pp. 347–362.

Müller-Trede, J., S. Sher, and C. R. M. McKenzie (2015). Transitivity in Context: A Rational Analysis of Intransitive Choice and Context-Sensitive Preference. *Decision 2*(4), 280–305.

Natenzon, P. (2019). Random Choice and Learning. *Journal of Political Economy 127*(1), 419–457.

Park, S., C. P. Davis-Stober, H. K. Snyder, W. Messner, and M. Regenwetter (2019). Cognitive Aging and Tests of Rationality. *The Spanish Journal of Psychology 22*(E57), 1–26.

Ranyard, R., H. Montgomery, E. Konstantinidis, and A. L. Taylor (2020). Intransitivity and Transitivity of Preferences: Dimensional Processing in Decision Making. *Decision 7*(4), 287–313.

Regenwetter, M., J. Dana, and C. P. Davis-Stober (2010). Testing Transitivity of Preferences on Two-Alternative Forced Choice Data. *Frontiers in Psychology 1*(148), 1–15.

Regenwetter, M., J. Dana, and C. P. Davis-Stober (2011). Transitivity of Preferences. *Psychological Review 118*(1), 42–56.

Savage, L. J. (1954). *The Foundations of Statistics*. New York: John Wiley & Sons.

Savage Jr., R. P. (1994). The Paradox of Nontransitive Dice. *American Mathematical Monthly 101*(5), 429–436.

Schmidt, U. and M. Stolpe (2011). Transitivity in Health Utility Measurement: An Experimental Analysis. *Health Economics Review 1*(1), 1–12.

Shafer, W. J. (1974). The Nontransitive Consumer. *Econometrica 42*, 913–919.

Shafir, S. (1994). Intransitivity of Preferences in Honey Bees: Support for Comparative Evaluation of Foraging Options. *Animal Behaviour 48*(1), 55–67.

Sopher, B. and G. Gigliotti (1993). Intransitive Cycles: Rational Choice or Random Error? An Answer Based on Estimation of Error Rates with Experimental Data. *Theory and Decision 35*(3), 311–336.

Starmer, C. (1999). Cycling with Rules of Thumb: An Experimental Test for a New Form of Non- Transitive Behaviour. *Theory and Decision 46*, 141–158.

Starmer, C. (2000). Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice Under Risk. *Journal of Economic Literature 38*(2), 332–382.

Starmer, C. and R. Sugden (1998). Testing Alternative Explanations of Cyclical Choices. *Economica 65*(259), 347–361.

Tversky, A. (1969). Intransitivity of Preferences. *Psychological Review 76*, 31–48.

Tversky, A., P. Slovic, and D. Kahneman (1990). The Causes of Preference Reversal. *American Economic Review 80*(1), 204–217.

Tversky, A. and R. H. Thaler (1990). Anomalies: Preference Reversals. *Journal of Economic Perspectives 4*(2), 201–211.

Waite, T. A. (2001). Intransitive Preferences in Hoarding Gray Jays (Perisoreus Canadensis). *Behavioral Ecology and Sociobiology 50*(2), 116–121.