Psychological Review
2011, Vol. 118, No. 1, 42–56

0033-295X/11/$12.00 DOI: 10.1037/a0021150

Transitivity of Preferences

Michel Regenwetter
University of Illinois at Urbana–Champaign

Jason Dana
University of Pennsylvania

Clintin P. Davis-Stober
University of Missouri—Columbia

Transitivity of preferences is a fundamental principle shared by most major contemporary rational, prescriptive, and descriptive models of decision making. To have transitive preferences, a person, group, or society that prefers choice option *x* to *y* and *y* to *z* must prefer *x* to *z*. Any claim of empirical violations of transitivity by individual decision makers requires evidence beyond a reasonable doubt. We discuss why unambiguous evidence is currently lacking and how to clarify the issue. In counterpoint to Tversky's (1969) seminal "Intransitivity of Preferences," we reconsider his data as well as those from more than 20 other studies of intransitive human or animal decision makers. We challenge the standard operationalizations of transitive preferences and discuss pervasive methodological problems in the collection, modeling, and analysis of relevant empirical data. For example, violations of weak stochastic transitivity do not imply violations of transitivity of preference. Building on past multidisciplinary work, we use parsimonious mixture models, where the space of permissible preference states is the family of (transitive) strict linear orders. We show that the data from many of the available studies designed to elicit intransitive choice are consistent with transitive preferences.

Keywords: rationality, transitivity, triangle inequality, utility theory, stochastic transitivity

Supplemental materials: http://dx.doi.org/10.1037/a0021150.supp

> Individuals "are not perfectly consistent in their choices. When faced with repeated choices between *x* and *y*, people often choose *x* in some instances and *y* in others." It seems "that the observed inconsistencies reflect inherent variability or momentary fluctuation in the evaluative process. This consideration suggests that preference should be defined in a probabilistic fashion." (Tversky, 1969, p. 31)

Michel Regenwetter, Department of Psychology, University of Illinois at Urbana–Champaign; Jason Dana, Department of Psychology, University of Pennsylvania; Clintin P. Davis-Stober, Department of Psychological Sciences, University of Missouri—Columbia.

This material is based upon work supported by the Air Force Office of Scientific Research, Cognition and Decision Program, under Award FA9550-05-1-0356 entitled "Testing Transitivity and Related Axioms of Preference for Individuals and Small Groups" (to Michel Regenwetter, principal investigator); by the National Institute of Mental Health, under Training Grant Award PHS 2 T32 MH014257 entitled "Quantitative Methods for Behavioral Research" (to Michel Regenwetter, principal investigator); and by the Decision, Risk and Management Science Program of the National Science Foundation, under Award SES 08-20009 entitled "A Quantitative Behavioral Framework for Individual and Social Choice" (to Michel Regenwetter, principal investigator). The final version of this article was written while Michel Regenwetter was a fellow of the Max Planck Institute for Human Development (ABC Group) and while Clintin P. Davis-Stober was the recipient of a Dissertation Completion Fellowship from the University of Illinois. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the funding agencies or the individuals who have provided comments or pointers. The human participants research was approved by the Institutional Review Board of the University of Illinois at Urbana–Champaign, under IRB 07520.

A large number of scholars have provided comments, input, and/or data at various times, including in presentations at the University of California at Irvine; the Universität Heidelberg, INSEAD; the Universität Mannheim; the Max Planck Institute Berlin; the University of Maryland; the University of Nijmegen; the University of Oklahoma; the University of Pennsylvania; the 2006 Annual Ward Edwards Bayesian Meeting; the 2006 Tucson Workshop on Decision Modeling and Behavior in Uncertain and Complex Environments; the 2006 DIMACS Workshop on Polyhedral Combinatorics of Random Utility; the 2006 and 2007 annual meetings of the Society for Mathematical Psychology; the 2006 and 2007 annual meetings of the European Mathematical Psychology Group; the 2006, 2007, and 2008 meetings of the Air Force Office of Scientific Research Cognition and Perception Program; the 2007 Institute for Operations Research and the Management Sciences Meeting; and the 2008 Society for Judgment and Decision Making Meeting.

Special thanks to Michael Birnbaum, Ulf Böckenholt, David Budescu, Jerome Busemeyer, Jean-Paul Doignon, Jean-Claude Falmagne, Klaus Fiedler, Samuel Fiorini, David Krantz, Graham Loomes, Duncan Luce, Tony Marley, John Miyamoto, and Peter Wakker (and many others) for helpful comments, pointers, and/or data. We are very grateful for the extensive help from graduate research assistants William Messner and Ying Guo. William Messner took care of the participant recruitment and the data collection. Both helped with data analysis and with the preparation of this article.

Correspondence concerning this article should be addressed to Michel Regenwetter, Department of Psychology, 603 East Daniel Street, University of Illinois at Urbana–Champaign, Champaign, IL 61822. E-mail: regenwet@illinois.edu

Jim is writing a dissertation on decision making and meets with his advisor three times a week. Each time, he offers his advisor a choice of two locations on campus. After a while, Jim notices that out of 30 times that he offered to meet in either the alumni center (A) or the business library (B), his advisor chose the alumni center 20 times. Out of the 30 times that Jim offered to meet in either the business library (B) or the coffee shop (C), his advisor preferred the business library 20 times. However, out of the 30 times that Jim offered to meet in either the alumni center (A) or the coffee shop (C), his advisor preferred to meet at the coffee shop 20 times. Jim questions the rationality of an advisor who two thirds of the time chooses A over B and two thirds of the time chooses B over C, but two thirds of the time chooses C over A, seemingly indicating intransitive preference. He considers switching to Bob's advisor, who, when given the same choices, chose A over B in 90% of cases, chose B over C in 90% of cases, and 70% of the time chose A over C.

Forlorn, Jim confronts his advisor, who gives him a surprisingly simple explanation for the apparent paradox. The two had met each Monday, Wednesday, and Friday when his advisor was on campus to teach, and she had chosen to meet closest to where she taught each time. On Mondays, she taught in the engineering building adjacent to the alumni center (A). On Wednesdays, she taught next to the business library (B), and on Fridays, she taught in the mathematics building across the street from the coffee shop (C). From the engineering building, the meeting locations are ranked ABC by proximity. From the classroom next to B, it is closer to C than to A. From the classroom near the coffee shop, the second closest venue is A. Hence, two of the three classrooms are closer to A than to B, two are closer to B than to C, and two are closer to C than to A. Jim's adviser had been choosing consistently according to a criterion—distance from current location—that cannot be cyclical.

Furthermore, she explains that Bob's advisor could not possibly be choosing in the same manner. If Bob's advisor is 90% of the time closer to A than B and 90% of the time closer to B than C, then it is mathematically impossible for him to be closer to C than A as often as 30% of the time even if he teaches class in a hot air balloon! Regardless of switching locations, he would have to choose A over C at least 80% of the time. The only person whose rationality can possibly be drawn into question is Bob's advisor.

This article discusses the challenges of disentangling variability in choices from structural inconsistency of preferences. As the Tversky epigraph suggests, when someone is faced with the same choice options repeatedly, he or she does not always choose the same way. As the epigraph also suggests, this may lead one to question whether the decision maker has consistent preferences. An important distinction must be drawn, however, between two possible meanings of inconsistent preference. Preferences could be inconsistent because they vary from time to time, that is, because the decision maker does not want the same thing at all times. We henceforth refer to this kind of inconsistency as variability. On the other hand, preferences could be inconsistent because they are logically incompatible with the assumptions (known as axioms) of expected utility theory (von Neumann & Morgenstern, 1947). Preferences that are inconsistent in this manner challenge the notion of human rationality and suggest the need for theories based on more psychologically descriptive assumptions to replace rational choice models.

Experimental research in burgeoning areas such as judgment and decision making and behavioral economics often claims to find the latter sort of inconsistency, rendering the axioms of expected utility theory descriptively inadequate. Such claims, however, are surprisingly difficult to substantiate properly, as we demonstrate in this article using the example of transitivity. Because decision axioms are stated in algebraic, deterministic terms, testing them involves bridging an open conceptual gap to intrinsically variable, probabilistic choice data (see Luce, 1995, 1997; see also Busemeyer & Townsend, 1993; Carbone & Hey, 2000; Harless & Camerer, 1994; Hey, 1995, 2005; Hey & Orme, 1994; Iverson & Falmagne, 1985; Loomes, 2005; Loomes & Sugden, 1995, for important related discussions). Bridging the gap presents a twofold challenge: (a) specifying a probabilistic model of the axiom (i.e., introducing variability) and (b) testing that model of the axiom using appropriate statistical methods (i.e., determining whether structural consistency has been significantly violated).

Our above example demonstrates how variability and structural inconsistency can be conflated. Suppose that we had modeled the variability in Jim's advisor's choices as reflecting true deterministic preference plus random error. This approach would imply the condition of *weak stochastic transitivity* (Block & Marschak, 1960; Luce & Suppes, 1965; Tversky, 1969), according to which, if A is chosen over B at least 50% of the time and B is chosen over C at least 50% of the time, then A must be chosen over C at least 50% of the time. Weak stochastic transitivity seems an intuitive bridge between a deterministic axiom and variable behavior, and indeed, Jim's intuition was that his advisor's preferences were intransitive. Yet, when we are able to reveal the mental process behind the choices, it is transitive at each and every instance (minimizing distance always rank orders the options), even if seemingly inconsistent at the aggregate level because it violates weak stochastic transitivity.

Our approach to modeling choice variability, the *mixture model*, avoids the erroneous conclusion that the advisor's preference is intransitive. The mixture model assumes that choices vary because the decision maker is in different mental states (analogous to different locations on campus) at different points in time. That is, the decision maker has some probability distribution over mental states and chooses x over y if and only if her or his current mental state is one in which she or he prefers x to y. Formally, as later stated in Equation 5 in the text,

$$\underbrace{P_{xy}}_{\substack{\text{overt probability of} \\ \textbf{choosing}\text{ x over y}}} = \underbrace{\sum_{\succ \in \Pi, x \succ y} P_{\succ}}_{\substack{\text{total probability of latent} \\ \text{mental states} \succ \text{in which} \\ \text{x is }\textbf{preferred}\text{ to y}}}.$$

To accommodate a two-alternative forced-choice (2AFC) task like the one above, only (transitive) linear orders over the choice options are allowable mental states.

Perhaps counterintuitively, this model already places a constraint on observed choice probabilities that is more restrictive than weak stochastic transitivity: Every distinct triple of choice probabilities must satisfy the *triangle inequalities*, as set out later in the text in Formula 4:

$$P_{xy} + P_{yz} - P_{xz} \leq 1.$$

We make no parametric assumptions about the probability distribution over allowable mental states, nor do we allow the flexibility of additional noise in the empirical choice data. Yet, when we analyze many sets of human and animal choice data using a powerful test of the highly restrictive triangle inequalities, we conclude that violations appear to occur only within a Type I error rate (≤5%). Hence, preferences appear to be generally transitive, even for stimuli that were sometimes designed to elicit intransitive behavior. The extensive literature on intransitivity should be reconsidered. More importantly, decision-making researchers should reconsider how frequently rational choice models are actually violated.

We focus specifically on transitivity for several reasons. First and foremost, transitivity is arguably the most fundamental axiom of rational choice (Bar-Hillel & Margalit, 1988). Informally, if an agent prefers x to y and y to z, then she or he must prefer x to z to be transitive. Nearly all normative, prescriptive, and even descriptive theories of choice imply transitivity (Kahneman & Tversky, 1979; Luce, 2000; Savage, 1954; von Neumann & Morgenstern, 1947). Indeed, real-valued utility functions logically require transitivity of preference, since the real numbers themselves are transitively ordered. Abandoning transitivity jeopardizes the very fundamental hypothetical construct of utility, and it questions nearly all theories that rely on this construct. On the other hand, should transitivity hold, then research in noncompensatory decision models (e.g., lexicographic heuristics) would lose one of its preeminent sources of empirical motivation and support.

The example of transitivity is further interesting because, on the surface, it would appear that suitable probabilistic models of transitivity have been developed (most notably, stochastic transitivity) and that testing these models is straightforward—both notions that we challenge later in this article. Finally, transitivity is interesting because it has been subject to intense scrutiny. After decades of research, there appears to be broad consensus that the axiom is empirically violated in human and animal decision makers (see, e.g., Brandstätter, Gigerenzer, & Hertwig, 2006; Gonzalez-Vallejo, 2002).

Axiom testing itself serves a purpose similar to that of the study of universal cognitive processes, such as attention or memory, even though it is grounded in a different research tradition. Just as the study of fundamental cognitive processes delineates what the cognitive system can or cannot do, axiom testing delineates what classes of theories can or cannot account for observed behavior. Furthermore, whether transitivity is violated may yield other important clues to the cognitive processes underlying choice behavior.

In this article, we review problems with past approaches, and we discuss the current (lack of) evidence for intransitivity of preference. A companion study (Regenwetter, Dana, & Davis-Stober, 2010) provided the in-depth technical discussion of how one should and how one should not proceed when testing transitivity of preferences.

Luce's Challenge: Deterministic Axioms Versus Probabilistic Data

The property of transitivity of preference says that if a person, group, or society prefers some choice option x to some choice option y and they also prefer y to z, then they furthermore prefer x to z. Formally, this is as follows:

DEFINITION. A *binary relation* $\succ$ on a set of choice alternatives $\mathcal{C}$ is a collection of ordered pairs of alternatives. It is standard to write such pairs as $x \succ y$ and to read the relationship as "x is (strictly) preferred to y." A binary relation $\succ$ on $\mathcal{C}$ satisfies the *axiom of transitivity* if and only if the following is true.

$$\text{Whenever } x \succ y \text{ and } y \succ z, \text{ then } x \succ z \text{ (for all choice options } x, y, z \text{ in } \mathcal{C}). \quad (1)$$

A binary relation is *intransitive* if it is not transitive. Because any given empirical study relies on a finite set of stimuli, we assume throughout that $\mathcal{C}$ is finite.

Any convincing test of transitivity must, as we discuss above, separate the issue of variability in overt choice behavior from the algebraic requirement that latent preferences are transitive. More generally, bridging such theory–data gaps is one of the most profound (and unresolved) challenges to empirical testing of decision theories (Luce, 1995, 1997). *Luce's twofold challenge* is to (a) recast a deterministic theory as a probabilistic model (or a hypothesis) and (b) properly test that probabilistic model of the theory (or the hypothesis) on available data.

Before we provide our solution to Luce's challenge, we review the broad range of approaches that the existing literature on (in)transitive individual preferences covers. We argue that all reported violations of transitivity of which we are aware have offered unconvincing solutions to Luce's challenge.

Existing Probabilistic Models for the Axiom of Transitivity of Preference

In meeting the first component of Luce's challenge—specifying a probabilistic model—some studies of intransitive preference have relied on *pattern counting* approaches. Typically, this involves counting the number of cyclical choice triples across many respondents, who each made every paired comparison once. This number is used to descriptively measure the *degree of intransitivity* of a respondent group in a given experimental condition (Bradbury & Nelson, 1974; Budescu & Weiss, 1987; Chen & Corter, 2006; Gonzalez-Vallejo, Bonazzi, & Shapiro, 1996; Lee, Amir, & Ariely, 2009; May, 1954; Mellers & Biagini, 1994; Mellers, Chang, Birnbaum, & Ordonez, 1992; Ranyard, 1977; Riechard, 1991; Sopher & Narramore, 2000; Treadwell, Kearney, & Davila, 2000; Tversky, 1969).

We agree that this measure can be interpreted as an indication of transitivity if there are zero cycles. When the degree of intransitivity is positive, however, it is unclear what this number can tell us about how close a group of decision makers is to being transitive. Figure 1 illustrates how different counting procedures could yield dramatically different assessments of how intransitive a binary relation is. One preference relation has the decision maker preferring octagon shapes with larger numbers to octagons with smaller numbers, except that Octagon 0 is preferred to Octagon 2. The other has the decision maker preferring triangles with larger numbers to triangles with smaller numbers, except that Triangle 0 is preferred to Triangle 101. Each of these relations is just one pairwise reversal away from the transitive ordering $>$ of the integers 0–101 (as indicated by the dotted arrows). Yet the pref-

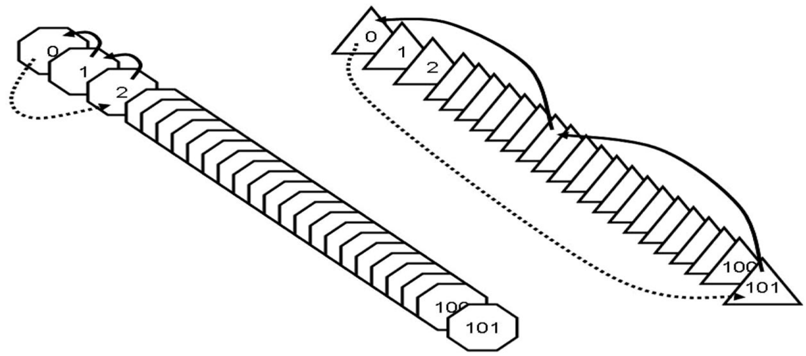

Figure 1. Two preference relations that each differ from the linear order $>$ on the integers 0, 1, . . . , 101, by one single ordered pair. The left-hand binary relation, where $0 > 2$ instead of $2 > 0$, has one cycle. The right-hand binary relation, where $0 > 101$ instead of $101 > 0$, has 100 cycles.

erence relation over octagons contains one cycle ($2 > 1$, $1 > 0$, $0 > 2$), while the preference relation over triangles contains 100 cycles ($101 > i$, $i > 0$, $0 > 101$, with $i = 1, 2, \ldots, 100$).

We also discuss the following surprising, but related, statistical fact in a later section: In an appropriate goodness-of-fit test, one group with few cyclical choice patterns could significantly violate a given model of transitivity while a second group with many more cyclical choice patterns does not. We illustrate later how degree of intransitivity is not monotonically related to the goodness of fit of our model.

Other approaches combine pattern counting with hypotheses, for example, that intransitive patterns occur significantly more often than expected by chance (Bradbury & Moscato, 1982; Bradbury & Nelson, 1974; Corstjens & Gautschi, 1983; Humphrey, 2001; Li, 2004; Peterson & Brown, 1998) or that a predicted cyclical pattern occurs significantly more often than its reverse (Kivetz & Simonson, 2000; Loomes, Starmer, & Sugden, 1991; Loomes & Taylor, 1992; Starmer, 1999). Regenwetter et al. (2010) showed, for example, that one approach leads to a self-contradiction while the other can confirm two incompatible theories on the same data.

When studying individual participants making repeated choices, researchers must accommodate variability over time. A simple approach for introducing probabilities here is to treat preference as deterministic and then attribute all variability to poor measurement, that is, error or noise in the data (for examples and/or discussions, see, e.g., Carbone & Hey, 2000; Harless & Camerer, 1994; Hey, 1995, 2005; Hey & Orme, 1994; Loomes, 2005). We consider this approach reasonable when the decision maker behaves in a highly consistent fashion. However, it is not uncommon for participants to choose one object over the other close to half the time. A theory that allows error rates approaching 50% is unsatisfying (see Loomes, 2005, for a related discussion).

In a related approach, Sopher and Gigliotti (1993) assumed that cyclical preferences have probability zero but that cycles occur in the choice data because paired-comparison responses are noisy and subject to errors. While they did not find support for intransitivity, they had to allow for rather large estimated error rates (e.g., exceeding 25%).

Undoubtedly, the most influential approach to Luce's challenge has been weak stochastic transitivity (Block & Marschak, 1960; Luce & Suppes, 1965). The leading article on this approach is Tversky (1969), of which we have found more than 600 citations in the past 20 years.

In his first and main experiment, Tversky (1969) used five gambles. He presented every participant with each of the 10 unordered pairs of gambles 20 times, separated by decoys. Tversky used a 2AFC paradigm: On each trial, the participant had to choose one gamble and was not allowed to express either indifference or lack of preference. Faced with the challenge of reconciling the structural axiom of transitivity with variability in the choice data, Tversky introduced probabilities as follows: x was said to be preferred to y only when it was chosen over y the majority of the time. Formally, he operationalized transitivity by the null hypothesis in the test (Formula 2) shown at the bottom of the page.

It is important to grasp the full complexity of these hypotheses. For instance, for three choice alternatives, as we vary the labels x, y, z, the null hypothesis considers six different inequality triples simultaneously. When dealing with five choice alternatives, there are 10 ways to select three distinct gambles x, y, z, with each selection leading to six different inequality triples. Thus, the null hypothesis considers 60 distinct inequality triples in that case. The alternative hypothesis considers two different cyclical scenarios for three gambles and 20 distinct cyclical scenarios for five gambles.

Weak stochastic transitivity, as a property of choice probabilities, is implied by a variety of models. For instance, decision field theory (Busemeyer & Townsend, 1993; Roe, Busemeyer, & Townsend, 2001) assumes that the utilities of objects at any point during the deliberation process are probabilistic, with a multivariate normal distribution. Other researchers who have considered Thurstonian random utility models include Böckenholt (1992b), Hey and Orme (1994), Roelofsma and Read (2000), and

$$\begin{cases} H_0: \text{For all distinct } x, y, z, \text{ if } (P_{xy} \geq 1/2) \text{ and } (P_{yz} \geq 1/2), \text{ then also } (P_{xz} \geq 1/2). \\ H_A: \text{There exist distinct } x, y, z \text{ with } (P_{xy} \geq 1/2), (P_{yz} \geq 1/2), \text{ and } (P_{xz} < 1/2). \end{cases} \quad (2)$$

Tsai and Böckenholt (2006). Such normally distributed random utility models imply weak stochastic transitivity as an immediate consequence (Halff, 1976).

Weak stochastic transitivity in conjunction with 2AFC holds if and only if majority aggregated preferences are transitive. As a consequence, one faces a conceptual complication with violations of weak stochastic transitivity: the *Condorcet paradox* of social choice theory (Condorcet, 1785). The paradox says that when transitive individual preferences are aggregated by majority rule, they can lead to cycles. Imagine three decision makers with the following preference orders (from best to worst): *xyz, yzx, zxy*. Even though each decision maker has a transitive rank ordering, the majority preference is cyclical: *x* preferred to *y* by a two-thirds majority, *y* preferred to *z* by two thirds, and *z* preferred to *x* by two thirds.

Tversky (1969) provided some evidence that the pairwise choices of an individual, when aggregated across trials, can generate a majority cycle, and he concluded that individual participants could behave intransitively. Yet the main reason why Condorcet's paradox is so notorious in social choice theory is because it shows that when aggregating any type of preferences (transitive or not) via majority rule, one can introduce cycles. In other words, following Tversky's own assessment that individuals' judgments and choices fluctuate, one can imagine an individual who has a transitive preference at every time point but violates weak stochastic transitivity at the aggregate (e.g., across time). Our introduction has illustrated such a situation, where a single individual rank ordered choice alternatives (transitively) by proximity to her current location but, because she was not always in the same location, generated choices that appeared intransitive in the aggregate.

The (mis)use of majority aggregated data is pervasive even in today's literature on individual decision making. For instance, Brandstätter et al. (2006) argued that the priority heuristic, which predicts intransitive behavior, outperformed all other theories they considered because it correctly predicted the largest proportion of observed majority choices (aggregated across participants). Yet, for $m \geq 3$, the majority choice in a sample of participants could have 100% agreement with the intransitive binary relation predicted by the priority heuristic even if none of the participants made their choices in accordance with the priority heuristic.

Aggregation paradoxes are not the only problem with using weak stochastic transitivity as a representation of transitivity. This model also implies additional properties that are substantially stronger than transitivity alone. When considering five choice alternatives, Regenwetter et al. (2010) showed that, for all practical purposes, weak stochastic transitivity reduces the aggregate preferences down to 120 linear orders. The number of (otherwise unconstrained) transitive relations is three orders of magnitude larger (see, e.g., Fiorini, 2001; Klaška, 1997). Rejecting weak stochastic transitivity on 2AFC data means essentially rejecting aggregate linear orders, not rejecting transitivity. On the other hand, if weak stochastic transitivity does hold, transitivity of majority aggregated preferences also holds.

Our main concern is with interpreting violations of stochastic transitivity as indicating violations of transitivity of preferences. Both the Condorcet paradox and the restrictiveness of stochastic transitivity show that such conclusions are unwarranted.

Finally, weak stochastic transitivity is theoretically weak in that it does not treat probability quantitatively. No distinction is made whether a person chooses *x* over *y* with probability 0.500001 or with probability one. Our preferred model, given in Equation 5 below, uses the binary choice probabilities on an absolute scale. Before we state that model, we also inspect the empirical paradigm itself.

The Problem of Two-Alternative Forced Choice

The most common empirical paradigm for testing transitivity of preferences, the 2AFC paradigm, itself raises problems for testing transitivity. The 2AFC task prohibits respondents from stating lack of preference or indifference between alternatives. Mathematically, this means that the 2AFC paradigm forces the axioms of asymmetry and completeness to hold in each observed binary choice datum (Regenwetter et al., 2010). This, in turn, implies that transitivity cannot be studied in isolation at the level of the observed choices. If one assumes that underlying preferences satisfy the axioms of completeness and asymmetry, then any transitive decision maker's preferences are characterized by a strict linear order. As we highlighted in our critique of weak stochastic transitivity, adding additional axioms is substantially stronger than just transitivity alone, especially when the number of choice alternatives is large.

For choices between gambles that involve complicated tradeoffs, we believe it is unrealistic to rule out either indifference or lack of preference. Forcing that additional structure, yet specifically claiming violations of transitivity, invites artifacts. While asymmetry is a normatively attractive property for strict preferences, we see no reason why a theory of rational choice would rule out indifference (say, among some similar, yet distinct, gambles with equal expected value). To accommodate the extra structure imposed on the empirical data, we later reanalyze the data under the null hypothesis of strict linear order preferences. Regenwetter and Davis-Stober (2010) proceeded to a different empirical paradigm, where they dropped the modeling assumption that preferences are linear orders.

In addition to the problems we have already discussed separately for the 2AFC task and for weak stochastic transitivity, problems arise from the combination of both, including further artifacts: Some pairs of stimuli in Tversky's (1969) study made indifference or high degrees of preference uncertainty plausible. This may have led respondents to make some pairwise choices with probability equal to or near one half. However, Tsetlin, Regenwetter, and Grofman (2003) showed in a social choice context that the latter behavior is particularly vulnerable to artificial majority cycles. When the choice probabilities are near one half, statistical tests of weak stochastic transitivity may have extremely high Type I error rates in small samples. This problem should not be taken lightly. For example, Kramer and Budescu (2005) illustrated the empirical prevalence of indifference in laboratory experiments.

We now proceed to review the model class that we believe offers the most natural solution to the first part of Luce's challenge.

Solving Luce's First Challenge: Mixture Models of Transitive Preference

Heyer and Niederée (1989, 1992), Regenwetter (1996), Regenwetter and Marley (2001), and Niederée and Heyer (1997) developed general tools to derive probabilistic generalizations from deterministic axiom systems. Ho, Regenwetter, Niederée, and Heyer (2005) demonstrated a successful application to the axiom of consequence monotonicity. This approach is not redundant with the approaches we have discussed earlier. Indeed, Loomes and Sugden (1995) concluded their study by pointing out that "future theoretical and empirical work should not regard stochastic specification as an 'optional add-on,' but rather as integral part of every theory which seeks to make predictions about decision making under risk and uncertainty" (p. 648). They and others showed that the same core theory can lead to opposite empirical predictions under competing stochastic specifications (e.g., Carbone & Hey, 2000; Hey, 1995, 2005; Hey & Orme, 1994; Loomes, 2005; Loomes & Sugden, 1995).

A mixture model of transitivity states that an axiom-consistent person's response at any time point originates from a transitive preference but that responses at different times need not be generated by the same transitive preference state. In particular, all responses could be generated by a single transitive cognitive process (e.g., "rank order venues by distance to current location"), and yet the overt responses at different times could differ from each other.

Formally, we assume that a person's responses come from a probability distribution over different possible transitive states. In the terminology of Loomes and Sugden (1995), this is a random preference model in that it takes a *core theory* (here, the axiom of transitivity) and considers all possible ways that the core theory can be satisfied. We use the term *mixture model* rather than the term *random preference model* to highlight the fact that the probability distribution underlying the observed patterns does not need to be uniform.

In the case of binary choices, again writing P_{xy} for the probability that a person chooses x over y and writing $\mathcal{T}$ for the collection of all transitive binary preference relations on $\mathcal{C}$, the mixture model states that

$$P_{xy} = \sum_{\substack{\succ \in \mathcal{T} \\ x \succ y}} P_{\succ}. \tag{3}$$

where $P_{\succ}$ is the probability that a person is in the transitive state of preference $\succ$ in $\mathcal{T}$. In words, the binary choice probability that x is chosen over y is the total (i.e., marginal) probability of all preference states in which x is preferred to y. Note that this model does not assume a 2AFC task, that is, $P_{xy} + P_{yx}$ need not be 1. Note also that $\mathcal{T}$ includes, for example, transitive, but incomplete, partial orders.

One way to formulate this as a psychological theory is to say that the space of preference states is $\mathcal{T}$ because any transitive preference is an allowable mental state. Each observation is collected at a randomly sampled preference state (from a possibly unknown distribution), either because the participant is randomly sampled or because the responses are separated by enough decoys to make the observations of a given respondent statistically independent of each other. Equation 3 implies that intransitive relations have probability zero.

Note that the probability distribution over $\mathcal{T}$ is not, in any way, constrained. This model does not imply weak stochastic transitivity of Formula 2, but it implies other constraints on binary choice probabilities, such as, for instance, the triangle inequalities (Marschak, 1960; Morrison, 1963; Niederée & Heyer, 1997), that is, for any distinct $x, y, z,$ in $\mathcal{C}$:

$$P_{xy} + P_{yz} - P_{xz} \leq 1. \tag{4}$$

The model stated in Equation 3 is closely related to the more restrictive classical *binary choice problem* (e.g., Marschak, 1960; Niederée & Heyer, 1997). In that problem, each decision maker is required to have strict linear order preferences (not just transitive relations), and $\mathcal{T}$ of Equation 3 is replaced by the collection of all strict linear orders over $\mathcal{C}$, which we denote by Π:

$$P_{xy} = \sum_{\substack{\succ \in \Pi \\ x \succ y}} P_{\succ}. \tag{5}$$

In words, the binary choice probability that x is chosen over y is the total (i.e., marginal) probability of all strict linear order preference relations in which x is preferred to y. This assumes a 2AFC task, that is, $P_{xy} + P_{yx} = 1$.

The study of the binary choice problem is intimately linked to the study of the *(strict) linear ordering polytope* (Bolotashvili, Kovalev, & Girlich, 1999; Cohen & Falmagne, 1990; Fiorini, 2001; Fishburn, 1992; Fishburn & Falmagne, 1989; Gilboa, 1990; Grötschel, Jünger, & Reinelt, 1985; Koppen, 1995; Suck, 1992), based on the fact that the permissible probabilities P_{xy} that satisfy Equation 5 form a *convex polytope.* It is well known that, for $m \leq 5$ (but not for larger m) and 2AFC, the triangle inequalities (Formula 4) are necessary and sufficient for the model representation (Equation 5). In other words, the triangle inequalities fully characterize the linear order model for up to five choice alternatives (see Regenwetter et al., 2010, for a thorough discussion).

Recall that Tversky (1969), as well as the intransitivity literature at large, used a 2AFC paradigm where respondents must choose either of two offered choice alternatives and that this forced the data to artificially satisfy the completeness and asymmetry axioms in each observed paired comparison. We pursue the case where the underlying model of preferences likewise assumes completeness and asymmetry. In the Discussion, we explain alternative routes (see also Regenwetter & Davis-Stober, 2008, 2010) that rely on different empirical paradigms and allow the participant to express indifference among choice alternatives.

To accommodate the use of a 2AFC paradigm, the canonical way to test whether data satisfy a mixture over transitive relations is to test the much more restrictive hypothesis that they lie in the (strict) linear ordering polytope, that is, test whether Equation 5 holds. Incidentally, the linear ordering polytope not only characterizes a model we consider theoretically superior but is also geometrically more restrictive (i.e., has a smaller volume) than the collection of polytopes defined by weak stochastic transitivity (Regenwetter et al., 2010). Thus, it is not only a more natural but also a more stringent test of transitivity. It is important to keep in mind, however, that violations of the linear ordering polytope, if found, are not necessarily due to violations of transitivity because strict linear orders are stronger than transitive relations. Most studies we have reviewed use five (or fewer) gambles, and thus, the triangle inequalities completely characterize the polytope in these cases.

We now turn our attention to Luce's second challenge, the use of appropriate statistical methods when testing models of transitivity.

Statistical Methods for Testing Models of the Axiom of Transitive Preference

Much of the literature on intransitive preferences does not even employ any statistical significance or quantitative goodness-of-fit tests at all but operates on a purely descriptive level (e.g., recently and prominently, Brandstätter et al., 2006). Those studies that have employed statistical testing often suffer from one or more of the problems we now review.

A common approach in the literature is to concentrate on only one cycle and to attempt to show that the data satisfy that cycle, often by pinning it against one single transitive preference relation, hence ignoring the quantifier in the axiom. For example, if there are five choice alternatives, there are 10 ways of choosing three objects, and each such triple can generate two cycles, hence yielding altogether 20 cyclical triples. Some researchers, rather than considering all possible cycles, only concentrated on one (e.g., Bateson, 2002; McNamara & Diwadkar, 1997; Schuck-Paim & Kacelnik, 2002; Shafir, 1994; Waite, 2001).

Intimately related to the problem of ignoring the quantifier is the problem of carrying out multiple binomial tests (e.g., McNamara & Diwadkar, 1997; Schuck-Paim & Kacelnik, 2002; Shafir, 1994; Waite, 2001). While the problem of ignoring the quantifier is a failure to consider all triples of alternatives jointly, the problem of multiple binomial tests is a failure to consider all pairs of alternatives jointly. For example, testing whether x is chosen over y with probability greater than one half and whether y is chosen over z with probability greater than one half through two separate binomial tests inflates Type I error. This problem is exacerbated when dealing with many pairs of alternatives. A Bonferroni correction (Hays, 1988) of the Type I error, in turn, causes the statistical power to deteriorate rapidly with increasing numbers of choice alternatives. The solution to this problem is to run a simultaneous test of all constraints. Such tests have recently become available.

The greatest threat to correct statistical testing in this context comes from boundary problems and the need for constrained inference. Many probabilistic generalizations of axioms lead to probabilistic models with inequality constraints on the parameters, similar to weak stochastic transitivity. This leads us to consider *constrained inference*, a domain of statistics where tremendous progress has been made in recent years. When the data themselves satisfy all of the inequality constraints, then there is not even a need for a statistical test because the model fits the data perfectly. When the data do not satisfy all constraints, however, one faces a so-called *boundary problem*. Here, the parameter point estimates will lie on the boundary of the parameter space, for example, on a face of a cube. Unfortunately, this means that certain requirements for standard likelihood theory are violated. As a consequence, for such models, the log-likelihood ratio statistic generally does not follow an asymptotic chi-square distribution.

Iverson and Falmagne (1985) demonstrated this problem for testing weak stochastic transitivity. In their reanalysis of Tversky (1969), only one of Tversky's violations of weak stochastic transitivity turned out to be statistically significant when analyzed with the correct asymptotic distribution. This fact does not appear to be well known: Compared to over 600 citations of Tversky in the past 20 years, we found Iverson and Falmagne cited fewer than 25 times. Of these, fewer than 10 (Birnbaum, 2004; Birnbaum & Gutierrez, 2007; Bouyssou & Pirlot, 2002; Gonzalez-Vallejo, 2002; Iverson, 2006; Karabatsos, 2006; Luce, 2005; Myung, Karabatsos, & Iverson, 2005; Tsai & Böckenholt, 2006) actually went as far as acknowledging explicitly that Iverson and Falmagne had falsified Tversky's analysis. While we have had to make many judgment calls on what precisely a given study claims, we have conservatively counted 110 post-1985 studies reporting that Tversky demonstrated violations of transitivity or of rational choice models generally (including such prominent works as Colman, 2003; Erev, 1998; Fishburn, 1986; Johnson-Laird & Shafir, 1993; Kivetz & Simonson, 2000; Luce, 1990; Mellers et al., 1992) The problem of incorrect statistical tests for weak stochastic transitivity is, of course, further compounded by the erroneous conclusion that a violation of weak stochastic transitivity demonstrates intransitive preferences.

Myung et al. (2005) developed a general Bayesian approach that is, in principle, directly applicable to these models, as long as all inequality constraints are explicitly known. Myung et al. revisited Tversky's (1969) study in a Bayesian model selection framework and concluded that, within a certain collection of candidate models, Tversky's conclusions outperformed the other models under consideration, thus reversing the substantive conclusions of Iverson and Falmagne (1985) from a Bayesian model selection perspective. We have several reservations with their conclusions. First, their list of candidate models did not include the class of mixture models that we endorse here. Second and more important, while a good fit of weak stochastic transitivity is evidence in favor of transitive aggregate preference and a good fit of the linear order model is evidence in favor of transitive individual preferences, the two models are not operating at the same level and thus do not actually stand in competition. This is compounded by the fact that violations of either of the two models would provide evidence only against linear order preferences (at either the aggregate or individual level), but not against transitivity per se.

An important implication of Iverson and Falmagne (1985) is the following: Consider the number of triples $a, b, c,$ such that, in a data set, a was chosen over b a majority of the time, b was chosen over c a majority of the time, and c was chosen over a a majority of the time. For $m > 3$, the number of such triples is not monotonically related to the goodness of fit of weak stochastic transitivity. Similar problems arise in other pattern counting approaches.

Solving Luce's Second Challenge: Statistical Test of the Linear Ordering Polytope

Recall that, for five alternatives, 20 distinct triangle inequalities completely characterize the linear ordering polytope, that is, the mixture model over linear orders. Recall also that one faces a constrained inference problem, where the log-likelihood ratio test statistic will fail to have an asymptotic chi-square distribution when the observed choice proportions lie outside the polytope. Instead, one needs to use a $\bar{\chi}^2$ (chi-bar-square) whose weights depend on the geometric structure of the polytope near the maximum-likelihood point estimate. Davis-Stober (2009) developed a method for finding that $\bar{\chi}^2$ distribution. We have used his method and refer the reader to Regenwetter et al. (2010) and to Davis-Stober (2009) for technical details.

In our data analysis, we assume that the observed choices form an independent and identically distributed (iid) random sample from an unknown distribution. This assumption raises the caveat that the experimenter must take precautions to design the experiment in a way that justifies the iid assumption.

A Reanalysis of Across-Participants Data

We applied the above methods in our reanalysis of across-participants data that were originally either published (Birnbaum, Patton, & Lott, 1999; Böckenholt, 1992a; Bradbury & Moscato, 1982; Bradbury & Nelson, 1974; Chen & Corter, 2006; Humphrey, 2001; Kirkpatrick, Rand, & Ryan, 2006; Kivetz & Simonson, 2000; Loomes, Starmer, & Sugden, 1989, 1991; Loomes & Taylor, 1992; May, 1954; Roelofsma & Read, 2000; Sopher & Gigliotti, 1993; Starmer, 1999; Starmer & Sugden, 1998) or made available to us (Birnbaum & Gutierrez, 2007; Lee et al., 2009).

Note that some of these studies had broader agendas than testing transitivity of preferences.

The online supplemental materials provide a table with the results for 107 across-participant data sets. A total of 12 studies yielded choice proportions lying outside the linear ordering polytope. Of these, only four were statistically significant violations at $\alpha = .05$ using the appropriate $\bar{\chi}^2$ distributions. The small proportion of violations is well within Type I error range.

In all of these 107 studies, it seems reasonable to treat the respondents as an iid sample from the population. However, some of these studies did not separate a given decision maker's paired comparisons by decoys, and some even allowed decision makers to revisit their prior choices and change them. With those studies, it may not be legitimate to treat all pairwise choices as resulting from iid sampling because individual decision makers may not have judged different pairs statistically independently of each other. This is why we do not dwell on these results. Rather, we move on to within-participant studies, such as Tversky's (1969) study, as well as later replication studies, including our own.

A Reanalysis of Tversky's and Others' Within-Participants Data

Table 1 shows our reanalysis of Tversky's (1969) Experiment 1. Out of 18 volunteers, eight students actually participated in the study, which used five different monetary gambles. Each participant provided 20 separate paired comparisons among all 10 possible gamble pairs. Tversky used decoys to avoid memory effects and thus attempted to ensure independent choices. Consider Respondent 1 from Table 1. This participant's choice proportions lie outside the linear ordering polytope and violate six different triangle inequalities. The log-likelihood ratio test statistic G^2 takes a value of 1.52. Using the methods developed by Davis-Stober (2009), we found that the local geometry at the point estimate yields the following asymptotic $\bar{\chi}^2$ distribution:

$$\bar{\chi}^2 = .08 + .41\chi_1^2 + .44\chi_2^2 + .07\chi_3^2 + .01\chi_4^2.$$

The test yields a p value of .34. In other words, the data may have landed outside the linear ordering polytope by sampling variability alone even if their underlying probabilities belonged to the polytope. As the table shows, six respondents generated choice proportions that violated the linear ordering polytope. Only one participant, Respondent 3, did so in a statistically significant fashion. Given that, in an initial screening task on a master set of 18 volunteers, the eight participants were cherry-picked for their alleged proneness to intransitivity, this raises the question whether the one violation may be within the scope of a Type I error. Note that we later demonstrate that our test has high statistical power.

If one assumes that the 10 persons who were screened out would not have generated significant violations, then Tversky's (1969) Experiment 1 data do not provide statistically compelling evidence for violations of the linear ordering polytope. Up to sampling variability, they are essentially consistent with the hypothesis that each respondent had a linear order preference state at each time point that he or she answered an item, with the preference relations being allowed to vary between responses. Tversky's data may be explainable by variability in choices without intransitivity in preferences. Recall also that linear orders are stronger than transitive relations. (For five objects, as here, there are more than 1,000 times as many transitive relations as linear orders.) The statistical evidence against transitivity is weak in this experiment, even though the study was custom designed to elicit violations.

Table 1 illustrates the danger of pattern counting. Respondent 1 violated six of the 20 triangle inequalities, and the test has a p value of .34. Yet Respondent 5 violated only a single triangle

Table 1
Reanalysis of the First Experiment in Tversky (1969) for the Eight Participants Who Participated in the Experiment (Out of 18 Who Entered the Prescreening)

Respondent	Number of triangle inequalities violated	Asymptotic distribution of G^2	G^2	p value
1	6	$.08 + .41\chi_1^2 + .44\chi_2^2 + .07\chi_3^2 + .01\chi_4^2$	1.52	.34
2	2	$.25 + .75\chi_1^2$	0.08	.59
3	7	$.05 + .16\chi_1^2 + .39\chi_2^2 + .34\chi_3^2 + .07\chi_4^2 + .01\chi_5^2$	**10.36**	**.01**
4	2	$.29 + .66\chi_1^2 + .05\chi_2^2$	0.91	.25
5	1	$.5 + .5\chi_1^2$	0.71	.20
6	5	$.05 + .14\chi_1^2 + .29\chi_2^2 + .33\chi_3^2 + .15\chi_4^2 + .04\chi_5^2 + .01\chi_6^2$	7.54	.05
7	0		0	
8	0		0	
9–18				

Note. For each respondent, we report how many triangle inequalities are violated by the choice proportions, the appropriate asymptotic $\bar{\chi}^2$-distribution, the log-likelihood ratio G^2 at the maximum-likelihood estimate, and the p value. Significant violations are marked in bold. Weights of the chi-bar distributions are rounded to two significant digits.

inequality but has a p value of .20. Even though Respondent 1 violated 6 times as many constraints, all these empirical violations amass less statistical significance than does the single violation of Respondent 5. This shows that the number of pattern violations (here, violated triangle inequalities) is not monotonically related to the p value of the full-fledged likelihood ratio test.

This nonmonotonicity may seem counterintuitive at first, but the following simple thought experiment may help to see why pattern counting can be so misleading. Suppose that a person flips 10 different coins, each 20 times. Even if none of them gives exactly 10 heads, as long as each of them yields pretty close to 10 heads, there is no reason to reject the null hypothesis that all 10 coins are fair. So, the fact that 10 out of 10 coins generated choice proportions different from one half does not imply that any coins are biased. On the other hand, suppose that even just one of these coins always came up heads in those 20 trials. This would be strong evidence against the claim that all 10 coins are fair, regardless of what the remaining 19 coins yielded. So, counting the number of coins that violated the requirement of coming up heads in 10 of 20 trials would be deceptive. To return to our analysis, Respondent 1 violated six out of 20 triangle inequalities, but these six violations are so minor that they (even jointly) do not provide significant evidence against the model.

We have not reanalyzed Tversky's (1969) Experiment 2 because it collected only three observations per gamble pair for each participant. This is not enough to use the asymptotic results of Davis-Stober (2009). In the online supplemental materials, we provide reanalyses of data from Montgomery (1977), Ranyard (1977), Tsai and Böckenholt (2006), and Waite (2001).The first two of these data sets used cherry-picked participants, and their sample sizes were so small that an asymptotic distribution may not be warranted. Tsai and Böckenholt used an extremely large sample size for each respondent, and Waite used a large sample size. These two studies yielded somewhat opposite conclusions, with the latter generating four significant violations in 12 birds, but the former generating none in five humans.

Experiment

Much of the data in the literature on testing transitivity comes with a variety of caveats: (a) The 107 across-participant data sets in the online supplemental materials provided only pooled data across participants; (b) Tversky (1969), Montgomery (1977), and Ranyard (1977) reported cherry-picked data only, and the latter two used extremely small sample sizes; and (c) just two prior studies provided individual respondent data, worked with a large sample size, and avoided cherry-picking of respondents, namely, Waite (2001) and Tsai and Böckenholt (2006).

We decided to run a study within the 2AFC paradigm, with individual respondent data, with large sample size, with stimuli that had a historic track record of allegedly generating intransitive behavior, and with no cherry-picking of subjects. We replicated Tversky's (1969) seminal study (without prescreening participants) by recruiting 18 human participants, using contemporary equivalents of Tversky's five gambles, and collecting a sample size of 20 repetitions per gamble pair per person. With this replication of Tversky's original study forming the core, we added two new sets of gambles, one with equal expected values and one with nonmonetary outcomes. In all three cases, the more attractive prizes came with smaller probabilities of winning.

The equal expected value gambles were meant to be even more conducive to intransitive behavior given that they involve even more difficult trade-offs. The nonmonetary gambles were designed to expand beyond the usual world of small-stakes monetary gambles, make numerical utility calculations difficult, and render the experiment more interesting to the respondents.

Method

Participants

Participants were 18 undergraduates (11 females, seven males) at the University of Illinois at Urbana–Champaign who responded to a campus advertisement for a paid experiment. Participants were not prescreened in any way. All participants gave informed consent before participating.

Procedures

As in Tversky's (1969) original experiment, participants made repeated choices over gambles presented as probability wheels with verbal descriptions of the outcomes, with the following modifications. Gambles were presented via computer interface rather than paper and pencil. Payoffs from Tversky's original gambles (Cash I) were converted to present-day dollar equivalents. In addition to these gambles, we used two new sets of gambles: A second set of monetary gambles (Cash II) with expected values equal to $8.80, and a set of nonmonetary gambles whose outcomes were gift certificates redeemable for a specific good, for example, 40 movie rentals. All five nonmonetary prizes had equal dollar values, a fact that we did not disclose to the participants. To establish the preference ordering among the gift certificates, each participant ranked them from most to least preferred before receiving instructions for the choice task, so that there was no strategic incentive to misstate one's preferences. The probabilities and payoffs for all gambles are summarized in Table 2, while Figures 2 and 3 depict examples of the computer display for choosing gambles.

Prior to beginning, participants were informed that they would be paid a base fee of $10.00 and that one of their choices would be randomly selected to be played for real at the experiment's conclusion. For the first 18 rounds of choice, participants were given gambles with randomly selected probabilities and outcomes drawn from one of the three choice sets. These rounds were used as training and were not considered in the final data analysis. During training, if a participant chose a stochastically dominated option, the software prompted him or her to notify the experimenter. In this way, participants could be monitored for any initial confusion about the task. The experimenter simply explained the task again without providing the reason for the prompt.

After training, participants chose between gamble pairs drawn sequentially from the Cash I set, Cash II set, noncash set, and a distractor after each such sequence of three trials. The order of this sequence was kept constant so that gambles from a given set were maximally separated from each other. The distractor gambles were included to reduce memory effects. The distractor gambles consisted of a win probability drawn from a uniform distribution and

Table 2
Cash and Noncash Gambles Used in Our Experiment

			Cash I		
Gamble	a	b	c	d	e
Probability of winning	7/24	8/24	9/24	10/24	11/24
Payoff	$28.00	$26.60	$25.20	$23.80	$22.40
			Cash II		
Gamble	a	b	c	d	e
Probability of winning	0.28	0.32	0.36	0.40	0.44
Payoff	$31.43	$27.5	$24.44	$22.00	$20.00
			Noncash		
Gamble	a	b	c	d	e
Probability of winning	9/50	10/50	11/50	12/50	13/50
			Noncash		
Gamble	f	g	h	i	j
Prizes	About 15 sandwiches	About 40 movie rentals	About 40 coffees	About 7 paperback books	About 4 music CDs

Note. Cash I is a computer-based replication of the stimuli used by Tversky (1969) in his Experiment 1, but using 2007 U.S. dollar equivalents.

an outcome from one of the three gamble sets in Table 2 with equal probability.

Participants continued until they had made 20 choices over each gamble pair in each set, thus making 818 choices. The pair presented in a given round was chosen randomly subject to the constraint that it had not been used in any of the last five trials from that gamble set and that neither one of the gambles had appeared in the previous trial from that set. The side of the screen on which a gamble appeared was also randomized.

The distractor items should have enhanced decision makers' attention but also counteracted memory effects, thus making the assumption of independence more realistic than a booklet format. First, several rounds of choice would have to have elapsed between any repetition of a gamble pairing. Second, because of the distractor sets, it appeared as if each gamble set was comprised of a large number of possible gambles, instead of only five. Finally, the distractor items were so diverse that they encouraged cognitive effort in every round. For example, in some distractor trials, one gamble had a substantially higher expected value, thus greatly discouraging respondents from choosing gambles by coin flip. On some distractor trials, one gamble stochastically dominated the other, thus giving incentives for vigilance. The rich set of distractor items also helped combat monotony in the decision task.

After making all choices, participants played their chosen gamble from a randomly selected trial by drawing a marble from an urn that replaced the probability wheel.

Results

Table 3 summarizes our findings for the 18 participants. Each decision maker had to carry out all 10 paired comparisons 20 times in each of the conditions marked Cash I, Cash II, and noncash. In Cash I, four out of 18 respondents generated choice proportions outside the linear ordering polytope, but only one of them (Respondent 16) violated the polytope significantly ($\alpha = .05$). In Cash II, which made the trade-off between outcomes and probability of winning even more difficult (all gambles had the same expected value), eight out of 18 participants generated choice proportions outside the polytope. However, again, only Respondent 16 violated the linear ordering polytope in a statistically significant fashion. In the noncash condition, we found no statistically significant violations at all. All in all, we found fewer participants and conditions

Figure 2. Example display of a Cash I paired-comparison stimulus.

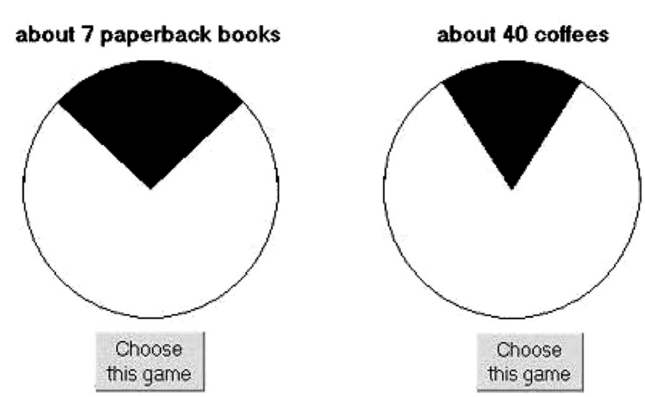

Figure 3. Example display of a noncash paired-comparison stimulus.

Table 3
Our Cross-Validation Study, With 18 Participants, and No Prescreening

	Cash I			Cash II			Noncash		
Respondent	G^2	*p* value	Degree of intransitivity	G^2	*p* value	Degree of intransitivity	G^2	*p* value	Degree of intransitivity
1	0		30	2.01	.29	51	0		8
2	0		15	0		3	2.87	.09	7
3	0		0	0		3	1.42	.31	2
4	3.77	.09	47	.10	.71	23	0		0
5	0		4	0		11	0		4
6	.36	.48	23	.08	.39	34	0		0
7	0		13	0		26	3.64	.08	4
8	0		0	0		2	0		13
9	0		39	0		21	0		24
10	0		6	.37	.55	10	0		2
11	0		1	1.42	.23	4	0		0
12	0		42	.000	.55	32	0		1
13	0		33	0		49	0		12
14	0		0	0		0	2.9	.18	2
15	0		27	0		27	0		0
16	**16.47**	**<.01**	29	**9.51**	**<.01**	46	1.43	.29	3
17	1.5	.17	48	0		17	0		9
18	0		37	.33	.48	55	.38	.52	6

Note. Each participant participated in all three scenarios, Cash I, Cash II, and Noncash. Significant violations of the linear ordering polytope are marked in bold (Respondent 16 in Cash I and Cash II).

with significant violations in our data than one would expect by Type I error.

We have seen in Table 1 that the number of violated triangle inequalities is not monotonically related to the *p* value of a quantitative test for the linear order mixture model. Similarly, Table 3 and Figure 4 document the nonmonotonic relationship between the most common form of pattern counting and the *p* value. For each respondent, we computed a version of what some authors call the

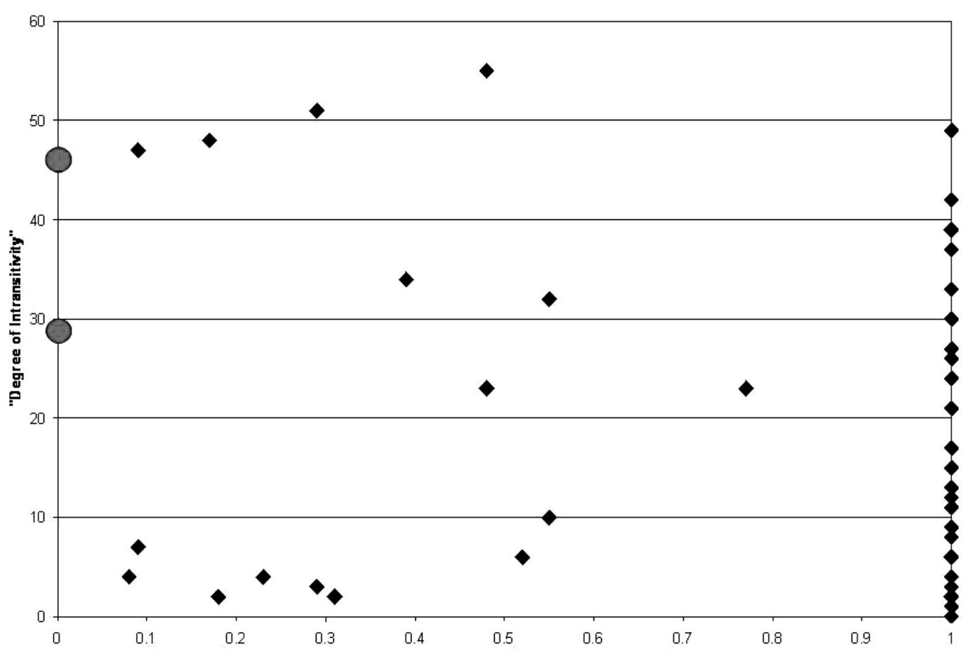

Figure 4. Pattern counting versus *p* value of quantitative test. The horizontal axis gives the *p* value of the quantitative fit for our 18 respondents across three gambles sets, with the two significant violations given on the far left as large circles and all perfect fits given on the far right. The vertical axis provides the number of cyclical triples in a pattern counting analysis, that is, what some authors call the degree of intransitivity.

degree of intransitivity as follows: First, we took the respondent's first binary choice for each gamble pair and combined these first choices into a first binary relation. Then, we took the respondent's second choice for each gamble pair and created a second binary relation. We proceeded the same way to construct 20 binary relations from each respondent. We then counted all intransitive triples across all 20 binary relations, separately for each respondent and each gamble set. As Table 3 and Figure 4 show, the two significant violations of the linear ordering polytope corresponded to degrees of intransitivity of 29 and 46, respectively. Twelve data sets that did not significantly violate the linear ordering polytope (including six perfect fits) generated a degree of intransitivity exceeding 29. The 52 data sets that did not significantly violate the polytope have degree of intransitivity pattern counting values that vary wildly between 0 and 55. This is particularly true for the 23 perfect fits of the linear ordering polytope, whose degree of intransitivity indices appear to be nearly uniformly distributed. Hence, the degree of intransitivity is undiagnostic of the fact that these data are compatible with a model that requires (transitive) linear order preferences with probability one.

Finally, we consider our statistical power to reject the mixture model when alternative models hold. The linear ordering polytope is a null hypothesis with nonzero volume in a high-dimensional space. There are infinitely many alternative hypotheses that can lie in infinitely many different directions in space. To tackle power in a general fashion, we wanted to avoid isolating any single specific alternative model. Instead, following an approach similar to that of Follman (1996), we considered 1,000 different alternative hypotheses. We selected these 1,000 points outside of the linear ordering polytope through uniform sampling from the outcome space. Each of these points forms a viable alternative hypothesis against which we can pin the mixture model. Using a simulation with a sample size of 20 repetitions per gamble pair to mimic Tversky's (1969) and our studies, we fit a data set drawn from each alternative hypothesis and correctly rejected the linear ordering polytope in 76.6% of those cases. Hence, we have high estimated average power. Interestingly, of those 1,000 data sets that we simulated by drawing from 1,000 alternative hypotheses, no single data set accidentally landed inside the linear ordering polytope. Recall from Table 3 that 23 out of 54 data sets lay inside the linear ordering polytope. This suggests to us that our findings are no statistical accident.

Conclusion and Discussion

Transitivity of preferences is a quintessential ingredient of nearly all normative, prescriptive, and descriptive theories of decision making. Almost any theory that uses utility functions implies or assumes transitivity. Yet the literature is peppered with reports of intransitive choice behavior. Some contemporary theory development in behavioral decision research, such as the development of noncompensatory decision models, is aimed at accommodating intransitive behavior (e.g., Brandstätter et al., 2006). We have provided a series of arguments challenging the common wisdom on intransitive preferences. In our view, there is little evidence of intransitivity.

Problems in the Existing Literature

As Luce (1995, 1997) discussed succinctly, several areas of psychology and of the decision sciences suffer from a conceptual and statistical disconnect between theory and data. Similarly, several scholars have extensively written about the need to integrate probabilistic components into deterministic decision theories and have cautioned about the conceptual challenges that arise in choosing a suitable stochastic specification (Carbone & Hey, 2000; Hey, 1995, 2005; Hey & Orme, 1994; Loomes, 2005; Loomes, Moffatt, & Sugden, 2002; Loomes & Sugden, 1995, 1998). Others (e.g., Iverson & Falmagne, 1985) have warned of nontrivial statistical pitfalls in psychological measurement. Nonetheless, these important warnings have generated limited impact, and few studies have even tackled Luce's challenge.

The algebraic axiom of transitivity (of preferences) is a prototypical and high-profile example of this problem. The conceptual problems associated with testing transitivity have received almost no attention (for the most notable exception, see Loomes & Sugden, 1995). A substantial portion of the literature on intransitive preferences completely fails to provide a probabilistic formulation of the algebraic axiom, and the vast majority of studies either ignore considerations of statistical inference or apply questionable statistical techniques.

In this article, we have discussed the most serious problems, namely, those associated with pattern counting, weak stochastic transitivity, the 2AFC paradigm, order-constrained inference, and cherry-picking. We have also provided what we consider the most careful empirical analysis of the existing evidence for intransitive preferences. We are aware of an additional handful of important, but somewhat less prominent, conceptual, mathematical, and statistical errors in the literature, which we have sketched only in passing. A technical companion article (Regenwetter et al., 2010) provided a complete in-depth analysis of the one dozen or so problems we have identified with the existing literature on intransitive preferences.

Mixture Models of Transitive Preference

Mixture models provide an elegant solution to Luce's first challenge. The basic idea is to take a core theory (e.g., transitivity, or the concept of a linear order) and require that theory to hold at each sample point in the probability space under consideration. This means that, in contrast to many other approaches, the mixture model requires the core theory to hold with probability one. We can identify our probability space here directly with the set of all strict linear orders on $\mathcal{C}$ endowed with some probability measure. The philosophy behind mixture models is furthermore to model variability substantively rather than treating it as a nuisance. Contrary to many approaches, the mixture model treats choice probabilities as absolute scale quantities. In the present article, we have focused on linear orders because this is the canonical model to handle the omnipresent 2AFC paradigm, in which decision makers are not allowed to express indifference among choice alternatives. Our model is equivalent to the linear ordering polytope.

Regarding Luce's second challenge, the order-constrained inference framework of Davis-Stober (2009) and Myung et al. (2005) interfaces with the mixture model naturally, as soon as a complete description of the relevant polytope is known. We have applied Davis-Stober's approach to the characterization of the linear ordering polytope via the triangle inequalities. While linear orders are substantially more restrictive than some of the other families of transitive binary relations one might consider, we have demonstrated that the linear order model can account for nearly all data sets available to us, most notably, three sources (Tsai & Böckenholt, 2006; Tversky, 1969; and our own data) of human data that collected large numbers of repetitions from individual participants. However, we believe that a more direct test of transitive preferences requires moving to different empirical paradigms.

Future research on intransitive preferences should take into account the considerable conceptual, mathematical, and statistical complexities and caveats that we have discussed. For example, 2AFC tasks should be avoided. Barring the problem of appropriate incentivization, researchers may consider ternary paired-comparison data, by which we mean decisions between pairs of choice alternatives with the option of expressing indifference between the two alternatives. This paradigm opens up the study of many convex polytopes (Regenwetter & Davis-Stober, 2008, 2010). For instance, Regenwetter and Davis-Stober (2010) investigated the weak order polytope for five choice alternatives. They found that this model also accommodates ternary paired-comparison choices.

Similar methods can be brought to bear on alternative mixture models, including mixtures of intransitive preferences. While our results here and the results of Regenwetter and Davis-Stober (2010) strongly suggest that preferences are transitive linear or weak orders, we cannot rule out that a mixture model of intransitive preferences could conceivably be found that also accounts for the same data. However, given that so many data sets have generated a perfect fit with our (very) parsimonious model, we have set an extremely high bar for competing models. Like elsewhere in science, the most parsimonious model that accounts for the empirical evidence should prevail until a better model is found.

Last but not least, we are not the first research team to consider stochastic specification in combination with intransitivity of preferences. This combination has previously received careful attention by Loomes, Starmer, and Sugden, whose extensive and seminal work in this area we have discussed in various places. Our conclusions and approach differ in the following main ways: (a) We have made a strong case against stochastic transitivity as a model of transitive preference in favor of pursuing mixture models. (b) There are many different possible mixture models for transitive preference (Regenwetter & Davis-Stober, 2008), and contrary to a prior cursory assessment by Loomes and Sugden (1995, p. 646), these mixture models are quite parsimonious. Moreover, most of them have never been studied empirically and many have hardly been studied mathematically. (c) We have emphasized convex geometric representations of mixture models via convex polytopes. (d) We have highlighted the fact that many of these stochastic specifications lead to order-constrained inference, for which appropriate statistical tools have only recently become available. (e) In other contexts, Loomes, Starmer, and/or Sugden have shown that mixture models can sometimes be so restrictive that they are readily rejected on empirical data (see, e.g., Loomes, 2005, for a discussion). For instance, this is typically the case when mixture models imply equality constraints among choice probabilities. An open question, to deal with this case, is how to create hybrids of mixture models and error models. We have not considered these here because they would only further reduce the already low number of statistically significant violations.

References

Bar-Hillel, M., & Margalit, A. (1988). How vicious are cycles of intransitive preference? *Theory and Decision, 24,* 119–145.

Bateson, M. (2002). Context-dependent foraging choices in risk-sensitive starlings. *Animal Behaviour, 64,* 251–260.

Birnbaum, M. H. (2004). Causes of Allais common consequence paradoxes: An experimental dissection. *Journal of Mathematical Psychology, 48,* 87–106.

Birnbaum, M. H., & Gutierrez, R. (2007). Testing for intransitivity of preferences predicted by a lexicographic semiorder. *Organizational Behavior and Human Decision Processes, 104,* 96–112.

Birnbaum, M. H., Patton, J. N., & Lott, M. K. (1999). Evidence against rank-dependent utility theories: Tests of cumulative independence, interval independence, stochastic dominance, and transitivity. *Organizational Behavior and Human Decision Processes, 77,* 44–83.

Block, H. D., & Marschak, J. (1960). Random orderings and stochastic theories of responses. In I. Olkin, S. Ghurye, H. Hoeffding, W. Madow, & H. Mann (Eds.), *Contributions to probability and statistics* (pp. 97–132). Stanford, CA: Stanford University Press.

Böckenholt, U. (1992a). Loglinear representation for multivariate choice data. *Mathematical Social Sciences, 23,* 235–250.

Böckenholt, U. (1992b). Thurstonian representation for partial ranking data. *British Journal of Mathematical & Statistical Psychology, 45,* 31–49.

Bolotashvili, G., Kovalev, M., & Girlich, E. (1999). New facets of the linear ordering polytope. *SIAM Journal on Discrete Mathematics, 12,* 326–336.

Bouyssou, D., & Pirlot, M. (2002). Nontransitive decomposable conjoint measurement. *Journal of Mathematical Psychology, 46,* 677–703.

Bradbury, H., & Moscato, M. (1982). Development of intransitivity of preference: Novelty and linear regularity. *Journal of Genetic Psychology, 140,* 265–281.

Bradbury, H., & Nelson, T. (1974). Transitivity and the patterns of children's preferences. *Developmental Psychology, 10,* 55–64.

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review, 113,* 409–432.

Budescu, D. V., & Weiss, W. (1987). Reflection of transitive and intransitive preferences: A test of prospect theory. *Organizational Behavior and Human Decision Processes, 39,* 184–202.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic–cognitive approach to decision making in an uncertain environment. *Psychological Review, 100,* 432–459.

Carbone, E., & Hey, J. D. (2000). Which error story is best? *Journal of Risk and Uncertainty, 20,* 161–176.

Chen, Y.-J., & Corter, J. E. (2006). When mixed options are preferred to mutiple-trial decisions. *Journal of Behavioral Decision Making, 19,* 17–42.

Cohen, M., & Falmagne, J.-C. (1990). Random utility representation of binary choice probabilities: A new class of necessary conditions. *Journal of Mathematical Psychology, 34,* 88–94.

Colman, A. (2003). Cooperation, psychological game theory, and limitations of rationality in social interaction. *Behavioral and Brain Sciences, 26,* 139–198.

Condorcet, M. (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix* [Essay on the application of the probabilistic analysis of majority vote decisions]. Paris, France: Imprimerie Royale.

Corstjens, M., & Gautschi, D. (1983). Conjoint analysis: A comparative analysis of specification tests for the utility function. *Management Science, 29,* 1393–1413.

Davis-Stober, C. P. (2009). Analysis of multinomial models under inequality constraints: Applications to measurement theory. *Journal of Mathematical Psychology, 53,* 1–13.

Erev, I. (1998). Signal detection by human observers: A cutoff reinforcement learning model of categorization decisions under uncertainty. *Psychological Review, 105,* 280–298.

Fiorini, S. (2001). Determining the automorphism group of the linear ordering polytope. *Discrete Applied Mathematics, 112,* 121–128.

Fishburn, P. C. (1986). Implicit mean value and certainty equivalence. *Econometrica, 54,* 1197–1205.

Fishburn, P. C. (1992). Induced binary probabilities and the linear ordering polytope: A status report. *Mathematical Social Sciences, 23,* 67–80.

Fishburn, P. C., & Falmagne, J.-C. (1989). Binary choice probabilities and rankings. *Economic Letters, 31,* 113–117.

Follman, D. (1996). A simple multivariate test for one-sided alternatives. *Journal of the American Statistical Association, 91,* 854–861.

Gilboa, I. (1990). A necessary but insufficient condition for the stochastic binary choice problem. *Journal of Mathematical Psychology, 34,* 371–392.

Gonzalez-Vallejo, C. (2002). Making trade-offs: A probabilistic and context-sensitive model of choice behavior. *Psychological Review, 109,* 137–154.

Gonzalez-Vallejo, C., Bonazzi, A., & Shapiro, A. J. (1996). Effects of vague probabilities and of vague payoffs on preference: A model comparison analysis. *Journal of Mathematical Psychology, 40,* 130–140.

Grötschel, M., Jünger, M., & Reinelt, G. (1985). Facets of the linear ordering polytope. *Mathematical Programming, 33,* 43–60.

Halff, H. M. (1976). Choice theories for differentially comparable alternatives. *Journal of Mathematical Psychology, 14,* 244–246.

Harless, D. W., & Camerer, C. F. (1994). The predictive value of generalized expected utility theories. *Econometrica, 62,* 1251–1289.

Hays, W. (1988). *Statistics* (4th ed.). New York, NY: Holt, Rinehart & Winston.

Hey, J. D. (1995). Experimental investigations of errors in decision making under risk. *European Economic Review, 39,* 633–640.

Hey, J. D. (2005). Why we should not be silent about noise. *Experimental Economics, 8,* 325–345.

Hey, J. D., & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica, 62,* 1291–1326.

Heyer, D., & Niederée, R. (1989). Elements of a model-theoretic framework for probabilistic measurement. In E. E. Roskam (Ed.), *Mathematical psychology in progress* (pp. 99–112). Berlin, Germany: Springer.

Heyer, D., & Niederée, R. (1992). Generalizing the concept of binary choice systems induced by rankings: One way of probabilizing deterministic measurement structures. *Mathematical Social Sciences, 23,* 31–44.

Ho, M.-H., Regenwetter, M., Niederée, R., & Heyer, D. (2005). An alternative perspective on von Winterfeldt et al.'s (1997) test of consequence monotonicity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 365–373.

Humphrey, S. (2001). Non-transitive choice: Event-splitting effects of framing effects? *Economica, 68,* 77–96.

Iverson, G. J. (2006). An essay on inequalities and order-restricted inference. *Journal of Mathematical Psychology, 50,* 215–219.

Iverson, G. J., & Falmagne, J.-C. (1985). Statistical issues in measurement. *Mathematical Social Sciences, 10,* 131–153.

Johnson-Laird, P., & Shafir, E. (1993). The interaction between reasoning and decision making: An introduction. *Cognition, 49,* 1–9.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263–291.

Karabatsos, G. (2006). Bayesian nonparametric model selection and model testing. *Journal of Mathematical Psychology, 50,* 123–148.

Kirkpatrick, M., Rand, A., & Ryan, M. (2006). Mate choice rules in animals. *Animal Behaviour, 71,* 1215–1225.

Kivetz, R., & Simonson, I. (2000). The effects of incomplete information on consumer choice. *Journal of Marketing Research, 37,* 427–448.

Klaška, J. (1997). Transitivity and partial order. *Mathematica Bohemica, 122,* 75–82.

Koppen, M. (1995). Random utility representation of binary choice probabilities: Critical graphs yielding critical necessary conditions. *Journal of Mathematical Psychology, 39,* 21–39.

Kramer, K. M., & Budescu, D. V. (2005). Exploring Ellsberg's paradox in vague-vague cases. In R. Zwick & A. Rapoport (Eds.), *Experimental business research* (Vol. 3, pp. 131–154). Norwell, MA: Kluwer Academic.

Lee, L., Amir, O., & Ariely, D. (2009). In search of *homo economicus*: Cognitive noise and the role of emotion in preference consistency. *Journal of Consumer Research, 36,* 173–187.

Li, S. (2004). A behavioral choice model when computational ability matters. *Applied Intelligence, 20,* 147–163.

Loomes, G. (2005). Modelling the stochastic component of behaviour in epxeriments: Some issues for the interpretation of the data. *Experimental Economics, 8,* 301–323.

Loomes, G., Moffatt, P. G., & Sugden, R. (2002). A microeconometric test of alternative stochastic theories of risky choice. *Journal of Risk and Uncertainty, 24,* 103–130.

Loomes, G., Starmer, C., & Sugden, R. (1989). Preference reversal: Information-processing effect or rational nontransitive choice? *Economic Journal, 9,* 140–151.

Loomes, G., Starmer, C., & Sugden, R. (1991). Observing violations of transitivity by experimental methods. *Econometrica, 59,* 425–439.

Loomes, G., & Sugden, R. (1995). Incorporating a stochastic element into decision theories. *European Economic Review, 39,* 641–648.

Loomes, G., & Sugden, R. (1998). Testing different stochastic specifications of risky choice. *Economica, 65,* 581–598.

Loomes, G., & Taylor, C. (1992). Non-transitive preferences over gains and losses. *Economic Journal, 102,* 357–365.

Luce, R. D. (1990). Rational versus plausible accounting equivalences in preference judgments. *Psychological Science, 1,* 225–234.

Luce, R. D. (1995). Four tensions concerning mathematical modeling in psychology. *Annual Review of Psychology, 46,* 1–26.

Luce, R. D. (1997). Several unresolved conceptual problems of mathematical psychology. *Journal of Mathematical Psychology, 41,* 79–87.

Luce, R. D. (2000). *Utility of gains and losses: Measurement-theoretical and experimental approaches.* Mahwah, NJ: Erlbaum.

Luce, R. D. (2005). Measurement analogies: Comparisons of behavioral and physical measures. *Psychometrika, 70,* 227–251.

Luce, R. D., & Suppes, P. (1965). Preference, utility and subjective probability. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 3, pp. 249–410). New York, NY: Wiley.

Marschak, J. (1960). Binary-choice constraints and random utility indicators. In K. J. Arrow, S. Karlin, & P. Suppes (Eds.), *Proceedings of the first Stanford Symposium on Mathematical Methods in the Social Sciences, 1959* (pp. 312–329). Stanford, CA: Stanford University Press.

May, K. (1954). Intransitivity, utility, and the aggregation of preference patterns. *Econometrica, 22,* 1–13.

McNamara, T., & Diwadkar, V. (1997). Symmetry and asymmetry of human spatial memory. *Cognitive Psychology, 34,* 160–190.

Mellers, B. A., & Biagini, K. (1994). Similarity and choice. *Psychological Review, 101,* 505–518.

Mellers, B. A., Chang, S., Birnbaum, M., & Ordonez, L. (1992). Preferences, prices, and ratings in risky decision making. *Journal of Experimental Psychology, 18,* 347–361.

Montgomery, H. (1977). A study of intransitive preferences using a think aloud procedure. In H. Jungerman & G. de Zeeuw (Eds.), *Decision-making and change in human affairs* (pp. 347–362), Dordrecht, the Netherlands: Reidel.

Morrison, H. W. (1963). Testable conditions for triads of paired comparison choices. *Psychometrika, 28,* 369–390.

Myung, J., Karabatsos, G., & Iverson, G. (2005). A Bayesian approach to testing decision making axioms. *Journal of Mathematical Psychology, 49,* 205–225.

Niederée, R., & Heyer, D. (1997). Generalized random utility models and the representational theory of measurement: A conceptual link. In A. A. J. Marley (Ed.), *Choice, decision and measurement: Essays in honor of R. Duncan Luce* (pp. 155–189). Mahwah, NJ: Erlbaum.

Peterson, G., & Brown, T. (1998). Economic valuation by the method of paired comparison, with emphasis on evaluation of the transitivity axiom. *Land Economics, 74,* 240–261.

Ranyard, R. H. (1977). Risky decisions which violate transitivity and double cancellation. *Acta Psychologica, 41,* 449–459.

Regenwetter, M. (1996). Random utility representations of finite *m*-ary relations. *Journal of Mathematical Psychology, 40,* 219–234.

Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2010). Testing transitivity of preferences on two-alternative forced choice data. *Frontiers in Quantitative Psychology and Measurement.* Advance online publication. doi:10.3389/fpsyg.2010.00148

Regenwetter, M., & Davis-Stober, C. P. (2008). There are many models of transitive preference: A tutorial review and current perspective. In T. Kugler, J. C. Smith, T. Connolly, & Y.-J. Son (Eds.), *Decision modeling and behavior in uncertain and complex environments* (pp. 99–124). New York, NY: Springer.

Regenwetter, M., & Davis-Stober, C. P. (2010). *Choice variability versus structural inconsistency of preferences.* Unpublished manuscript.

Regenwetter, M., & Marley, A. A. J. (2001). Random relations, random utilities, and random functions. *Journal of Mathematical Psychology, 45,* 864–912.

Riechard, D. (1991). Intransitivity of paired comparisons related to gender and community socioeconomic setting. *Journal of Experimental Education, 59,* 197–205.

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision making. *Psychological Review, 108,* 370–392.

Roelofsma, P., & Read, D. (2000). Intransitive intertemporal choice. *Journal of Behavioral Decision Making, 13,* 161–177.

Savage, L. J. (1954). *The foundations of statistics.* New York, NY: Wiley.

Schuck-Paim, C., & Kacelnik, A. (2002). Rationality in risk-sensitive foraging choices by starlings. *Animal Behaviour, 64,* 869–879.

Shafir, S. (1994). Intransitivity of preferences in honey bees: Support for "comparative" evaluation of foraging options. *Animal Behaviour, 48,* 55–67.

Sopher, B., & Gigliotti, G. (1993). Intransitive cycles: Rational choice or random error? An answer based on estimation of error rates with experimental data. *Theory and Decision, 35,* 311–336.

Sopher, B., & Narramore, J. M. (2000). Stochastic choice and consistency in decision making under risk: An experimental study. *Theory and Decision, 48,* 323–350.

Starmer, C. (1999). Cycling with rules of thumb: An experimental test for a new form of non-transitive behavior. *Theory and Decision, 46,* 139–158.

Starmer, C., & Sugden, R. (1998). Testing alternative explanations of cyclical choices. *Economica, 65,* 347–361.

Suck, R. (1992). Geometric and combinatorial properties of the polytope of binary choice probabilities. *Mathematical Social Sciences, 23,* 81–102.

Treadwell, J., Kearney, D., & Davila, M. (2000). Health profile preferences of Hepatitis C patients. *Digestive Diseases and Sciences, 45,* 345–350.

Tsai, R.-C., & Böckenholt, U. (2006). Modelling intransitive preferences: A random-effects approach. *Journal of Mathematical Psychology, 50,* 1–14.

Tsetlin, I., Regenwetter, M., & Grofman, B. (2003). The impartial culture maximizes the probability of majority cycles. *Social Choice and Welfare, 21,* 387–398.

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

von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior.* Princeton, NJ: Princeton University Press.

Waite, T. A. (2001). Intransitive preferences in hoarding gray jays. *Behavioral Ecology and Sociobiology, 50,* 116–121.

Received March 15, 2008
Revision received May 11, 2010
Accepted May 15, 2010 ■

Correction to Pleskac and Busemeyer (2010)

In the article "Two-Stage Dynamic Signal Detection: A Theory of Choice, Decision Time, and Confidence" by Timothy J. Pleskac and Jerome R. Busemeyer (*Psychological Review, 117,* 864–901), the name of the philosopher Charles Peirce was misspelled throughout as Charles Pierce.

DOI: 10.1037/a0022399