Article:

Ranyard, R, Montgomery, H, Konstantinidis, E orcid.org/0000-0002-4782-0749 et al. (1 more author) (2020) Intransitivity and Transitivity of Preferences: Dimensional Processing in Decision Making. Decision, 7 (4). pp. 287-313. ISSN 2325-9965

https://doi.org/10.1037/dec0000139

Intransitivity and Transitivity of Preferences: Dimensional Processing in Decision Making

Rob Ranyard

University of Leeds, UK

Henry Montgomery

Uppsala University, Sweden

Emmanouil Konstantinidis

Warwick University, UK

and Andrea L. Taylor

University of Leeds, UK

Author Note

Rob Ranyard and Andrea L. Taylor, Centre for Decision Research, Leeds University Business School, UK.

Henry Montgomery, Department of Psychology, Uppsala University, Sweden.

Emmanouil Konstantinidis, Department of Psychology, Warwick University, UK.

We are grateful for comments and feedback on earlier drafts from Peter Juslin, Wändi Bruine de Bruin and Ola Svenson. Our thanks are also due to Daniel Cavagnaro and to Tobias Kalenscher for providing us with the data from their studies.

Correspondence concerning this article should be addressed to Rob Ranyard, Centre for Decision Research, Leeds University Business School, Moorland Rd, Leeds, West Yorkshire, LS6 1AN, UK. email: r.ranyard@leeds.ac.uk.

Abstract

Transitive preference, i.e., if you prefer apples to bananas and bananas to cherries, you also prefer apples to cherries, is a basic property of some influential rational choice models. Contrary to this, Tversky, in his seminal 1969 article, presented evidence of intransitive preferences in two contexts, one of which involved choices between simple monetary lotteries. Whilst early replications corroborated his findings, more recent research cast doubt on the strength of evidence of intransitive preferences in this task. Here, from Tversky's extended additive difference model we develop a simplified additive difference (SAD) model which corresponds to alternative dimensional processing strategies. This predicts transitive or intransitive preferences, depending on its parameter values. We review six replications of Tversky's lottery task and fit variants of the model to the choice data. We estimate the SAD model's parameters for each individual data set using maximum likelihood estimation, examine the goodness of fit of the model and use likelihood ratio tests to evaluate specific variants. The model has a very good fit to most individual choice data sets reviewed, with many predictably violating weak stochastic transitivity. We also find that many transitive patterns correspond to the application of simple, one-dimensional 'take the best' heuristics. The findings support the view that human decision making is often based on dimensional processing in such a way that evaluations of decision alternatives are relative to the set under consideration, resulting in intransitivity of preferences.

Keywords: Risky choice; intransitivity; transitivity; dimensional processing; additive difference model

Intransitivity and Transitivity of Preferences: Dimensional Processing in Decision Making

Suppose someone prefers apples to bananas and bananas to cherries. Their preferences will be transitive if they also prefer apples to cherries, but intransitive if they prefer cherries to apples. Let us represent the strict preference relation “is preferred to” by the symbol $\succ$ and the above alternatives by their first letter. With this notation, a transitive cycle of preferences is one in which a $\succ$ b and b $\succ$ c implies a $\succ$ c, while the intransitive cycle concludes with the opposite, i.e., a $\succ$ b and b $\succ$ c but c $\succ$ a. In order to include indifference, such as liking apples and bananas equally, preferences are more usually described in terms of the relation “is preferred or indifferent to”, denoted by a $\succcurlyeq$ b. Transitivity is thus formally defined algebraically as follows.

(1) For all a, b, c in a set of alternatives, a $\succcurlyeq$ b and b $\succcurlyeq$ c implies a $\succcurlyeq$ c.

Testing the transitivity of preferences is important, not least because it is a necessary condition for the validity of utility theory as a description of human preference, even at an ordinal level (Edwards, 1954; von Neumann & Morgenstern, 1947). However, there are some basic aspects of preference that make its empirical testing a tricky problem; notably, preferences change over time or with context. Furthermore, even within a short time frame and in a fixed context, preferential choices can be inconsistent. That is, even if someone has a broad preference for apples over bananas, he or she will sometimes choose a banana over an apple. This kind of variability is conventionally accommodated by characterizing choice as probabilistic (or stochastic), with the probability of choosing a over b denoted by p(a,b).

In his landmark investigation, Intransitivity of Preferences, Tversky (1969) minimized the potential for changes in preference by investigating people’s choices within a relatively short time frame in experiments that, as far as possible, controlled the context. In addition, he accommodated inconsistency by substituting the algebraic relation, a $\succcurlyeq$ b, with the probabilistic relation, $p(a,b) \geq .5$. He then investigated the following property, known as weak stochastic transitivity (WST).

(2) For all a, b, c from a set of alternatives,

$$p(a,b) \geq .5 \text{ and } p(b,c) \geq .5 \text{ implies } p(a,c) \geq .5.$$

Regenwetter, Dana and Davis-Stober (2010, p. 6) explain that WST is equivalent to the weak utility model (Block and Marschak, 1960; Luce and Suppes, 1965), whereby choice probability is related to a utility function, U, as follows:

$$U(a) \geq U(b) \leftrightarrow p(a,b) \geq .5 \leftrightarrow a \succcurlyeq b.$$

Testing WST seems to be a most elegant formulation of the research question of whether decision makers' preferences are transitive or intransitive. After all, if preferential choice fails to satisfy even this weakest of probabilistic realizations of the transitivity principle, then surely this calls into question the descriptive validity of any decision theory that assumes it, including utility theories. In fact, Tversky's main empirical finding was that across two experiments the choice patterns of several individuals did significantly violate WST.

More recently, Regenwetter, Dana and Davis-Stober (2010, 2011) developed an alternative stochastic specification of transitive preferences, the mixture model, which requires the following alternative restriction on choice probabilities, known as the triangle inequalities (TI) condition.

(3) For all a, b, c from a set of alternatives,

$$p(a,b) + p(b,c) - p(a,c) \leq 1.$$

Regenwetter et al. (2010) explain that their mixture model is equivalent to the random utility model (Block & Marschak, 1960) and argue that testing the TI condition is an acceptable alternative test of transitive preferences. Based on their tests of TI with previously collected data, as well as on new data that they presented, Regenwetter et al. (2011) concluded that "unambiguous evidence is currently lacking" of "empirical evidence of intransitivity by individual decision makers" (p. 42). In this paper we aim to shed light on this issue, which is at the core of human decision making, by further exploring Tversky's theoretical insights, which were based on the notion of dimensional processing.

Tversky's (1969) important theoretical contribution was the development of two plausible decision models that predict intransitive choice in some circumstances: the additive difference model and the lexicographic semiorder (LS) heuristic. Both are based on dimensional processing: the comparison of advantages and disadvantages across choice alternatives within the various dimensions on which alternatives differ. If people do make decisions in such ways, the implications for the psychology of decision making are fundamental. On the one hand, the classes of theories related to utility theory, including prospect theory (Kahneman & Tversky, 1979), assume that each alternative is evaluated independently of the others in the available choice set. In contrast, dimensional processing models such as Tversky's assume that the evaluation of the each alternative is relative, and depends on which other alternatives are under consideration. Such relative evaluative processes would constitute a basic aspect of bounded rationality (Simon, 1956).

The analysis and review we present here aims to build on Tversky's (1969) pioneering work in ways that have previously been neglected. As well as adopting a process perspective, our treatment has several distinctive features, compared to previous treatments (e.g., Cavagnaro & Davis-Stober, 2014). First, previous research on this issue has tended to neglect descriptions of choice behavior. Rather, the focus has been on how well theoretical models fit choice behavior data. As Tukey (1980) has argued, however, exploring and describing data is just as important as model testing. In this spirit, we present an exploratory descriptive analysis of choice patterns in Tversky's (1969) lottery context. We see this as a crucial component in developing a process account of intransitive and transitive preferences when participants have complete information. (It should be noted that in contrast to the context investigated by Müller-Trede, Sher, and McKenzie (2015), participants in the studies reviewed here had full information about the attributes of alternatives relevant to their decision). Second, previous treatments proposed alternative models but did not evaluate the goodness of fit of Tversky's additive difference model to replications of his lottery task[1]. Here we do this by developing a two-parameter, simplified additive difference (SAD) model, and describe the relationship between the model's parameters and the transitivity or intransitivity of choice in Tversky's lottery choice task. That is, we identify the conditions under which the two-parameter SAD model predicts that WST will be satisfied or violated in this task. We further show how different specifications of the model correspond to different dimension-based decision strategies and heuristics: this draws on the assumption that decision makers have a repertoire of decision strategies from which they can select (Payne, 1982; Payne, Bettman & Johnson, 1993; Svenson, 1979). Process-tracing evidence, for example, from think aloud protocols (Montgomery, 1977), eye-tracking (Glöckner & Herbold, 2011), and information acquisition monitoring (Pachur, Hertwig & Wolkewitz, 2014)) has shown that across-gamble, dimension-based, as well as within-gamble, cognitive processing has a role in decisions between simple monetary gambles. As well as Tversky's (1969) dimension-based models we consider one-dimensional, 'take the best' heuristics

[1] The goodness of fit to some data sets of a stochastic version of Tversky's other model predicting intransitive preferences, the lexicographic semiorder, has been tested (Davis-Stober, Brown & Cavagnaro, 2015; Regenwetter, Dana, Davis-Stober & Guo, 2011).

(Gigerenzer & Goldstein, 1999), since some participants may resort to choice strategies requiring less cognitive effort when faced with the demands of a long series of similar choice problems.

Our main contribution, then, is that, based on Tversky's (1969) extended additive difference model, we develop a simplified additive difference (SAD) model, which allows for a quantitative description of how individuals vary with respect to dimensional processing strategies. Most importantly, the SAD model specifies boundaries for when violations of WST will occur for an individual depending on his or her dimensional processing strategy. Using this framework we review and reanalyze previous choice data from seven replications of Tversky's lottery choice experiment: Cavagnaro & Davis-Stober, (2014), Kalenscher et al. (2010), Montgomery (1977), Ranyard (1977), Regenwetter et al. (2011), Tsai and Böckenholt (2006) and Tversky (1969). We are aware that Regenwetter et al. (2010) have argued that WST is not a legitimate criterion of stochastic transitivity. However, we retain Tversky's original, and still widely recognized, WST criterion (e.g. Oliveira, Zehavi, & Davidov, 2018), because WST is a transparent and clearly defined formulation of stochastic transitivity such that: "a person's preference … is [stochastically] transitive if their majority choices (over repeated trials) are transitive" (Regenwetter et al., 2010, p. 5). Nevertheless, we also consider Regenwetter et al.'s TI alternative when comparing alternative analyses.

Our reanalysis involves a descriptive and a model fitting stage in which we estimate the SAD model's parameters for each individual data set, using the maximum likelihood estimation (MLE) method. We then test its goodness of fit relative to the alternative models described earlier and apply likelihood ratio tests to classify individual choice data by decision strategy type, and whether they are transitive or intransitive with respect to WST. Next we compare our findings to previous analyses that were from the perspective of transitive models. Our aim is to establish, once and for all, whether Tversky's (1969) original findings of intransitive preferences with respect to WST are robust and replicable, and whether his contribution to theory, with respect to dimensional processing in decision making, is valid. We conclude with some caveats and some avenues for future research

Tversky's dimensional processing models

Tversky (1969) argued that intransitive preferences might occur when people construe decision alternatives as varying on two or more dimensions and process information across dimensions rather than within alternatives. He considered two cases of dimensional decision representations: (1) choice between simple monetary lotteries with the structure win s dollars with probability p, otherwise win zero (dimensions S and P); and (2) choice between job applicants varying in three important dimensions that were labelled intellectual ability, emotional stability and social facility. His first example of a model that might result in intransitive preferences, the LS heuristic, processes only some of the available information. The second, the additive difference model, fully utilizes all available information by evaluating and comparing all dimension differences.

The lexicographic semiorder (LS) heuristic

With the LS heuristic, information is processed as follows:

> If the difference between the two alternatives on dimension I is (strictly) greater than [a threshold value,] ε, choose the value that has the higher value on dimension I. If the difference between the alternatives is less than or equal to ε, choose the alternative that has the higher value on dimension II (Tversky, 1969, p32).

In his first study, Tversky investigated choices between pairs from a set of simple lotteries as described above, with the values of P and S shown on the left of Table 1. It can be seen that adjacent lotteries in the sequence a, b, c, etc. have small differences on each dimension, whereas S and P differences between non-adjacent lotteries are greater. Suppose the probability of winning (P) is dimension I and the winning amount (S) is dimension II, the probability difference threshold, $\varepsilon = 2/24$ and choice inconsistency, or error rate, is fixed at 20%[2]. It should be noted that the priority heuristic in this context is the LS heuristic with $\varepsilon = 1/10$ (Brandstätter, Gigerenzer & Hertwig, 2006). The LS heuristic applied in this way to choices between pairs of alternatives from Tversky's lottery set could lead to the choice proportions shown in the top left panel of Table 1. This shows how the lottery with the higher S value is chosen most often when the P difference is less than or equal to the threshold, whereas that with the higher P value is chosen more often when it is greater. It can be seen how WST is violated for several cycles of choice in this case, for example choice proportions for (a,b), (b,c), (c,d) and (d,e) are all greater than .5, whereas that for (a,e) is less than .5. In contrast, the top right panel of the table illustrates choice proportions for an LS heuristic in which S is dimension I, P is dimension II, and the threshold is $\varepsilon = \$0.25$ on the S dimension. This leads to violations of WST in the opposite direction to the first case. It is notable that in both cases there is a sharp switch in preferences for either the higher S or the higher P as differences change.

It should be noted that the way the WST might be violated in this lottery set is different for different values of the LS heuristic threshold. If the probability difference threshold, $\varepsilon = 3/24$,

[2] Although a 20% error rate is arbitrary, it represents a relatively consistent but variable application of the heuristic, taking heed of Regenwetter et al.'s comment that "a theory that allows error rates to approach 50% is unsatisfying" (2011, p.45).

then only the cycle of five choice pairs are predicted to violate WST, i.e., p(a,b), p(b,c), p(c,d), p(d,e) ≥ .5, but p(a,e) < .5. However, if the threshold were 2/24, as illustrated in the top left panel of Table 1, the cycles of four pairs would also violate WST, i.e.: p(a,b), p(b,c), p(c,d) ≥ .5, but p(a,d) < .5; and p(b,c), p(c,d), (p(d,e) ≥ .5, but p(b,e) < .5. Finally, if the threshold were 1/24, the cycles of three adjacent pairs would also be predicted to violate WST.

--- Table 1 in here ---

The additive difference model

The additive difference model assumes that decisions are made by evaluating the subjective difference on each dimension and weighing additively those differences favoring one alternative (its advantages) against those not favoring it (its disadvantages). Like utility models, this decision rule is compensatory in that all information is fully utilized in trade-offs between advantages and disadvantages. In the case of choosing from a pair of two-dimensional lotteries, the model reduces to evaluating and comparing the P advantage with the S disadvantage, or vice-versa. For example, in Montgomery's (1977) replication of Tversky's (1969) lottery study one of his participants said: "I'll take the alternative B because the decrease in the chances of winning does not correspond to the increase in payoff" (p. 352).

Tversky's (1969) additive difference model (algebraic version)

The following account of the additive difference model closely follows Tversky (1969).

Let $A = A_1 \times A_2 \ldots\ldots \times A_n$ be a set of multidimensional alternatives with elements of the form:

$x = (x_1, \ldots\ldots, x_n)$, $y = (y_1, \ldots\ldots, y_n)$, where x_i and y_i ($i = 1, \ldots, n$) are the values of alternatives x and y on dimension i.

There exist scales u_1, u_2, ….. , u_n defined on A_1, A_2, ….. A_n respectively such that $u_i(x_i)$, $u_i(y_i)$ are the subjective values of the ith component of alternatives x and y respectively.

Then $\delta_i = u_i(x_i) - u_i(y_i)$ is defined as the difference between the subjective values of x and y on the ith dimension. A difference function, $\varphi_i(\delta_i)$, is applied to this which determines the subjective difference between x and y on the ith dimension. This can be viewed as the advantage or disadvantage of x over y on dimension i. Let $\varphi_i(-\delta_i) = -\varphi_i(\delta_i)$.

A preference structure satisfies the additive difference model if there exist real-valued functions u_1, u_2, ….. , u_n and continuous difference functions, $\varphi_1(\delta_1)$, …. $\varphi_n(\delta_n)$ such that:

(1) $x \succcurlyeq y$ if and only if $osd(x,y) = \sum_{i=1}^{n} \varphi_i[u_i(x_i) - u_i(y_i)] \geq 0$

where osd(x,y) is the overall subjective difference between x and y,

and $\varphi_i(-\delta_i) = -\varphi_i(\delta_i)$ for all i.

Tversky proved that under the additive difference model preference is transitive if and only if the following conditions apply, otherwise preference may be intransitive: (1) for all $n \geq 3$, transitivity holds if and only if all dimension difference functions are linear; (2) for $n = 2$, transitivity holds if and only if $\varphi_1(\delta) = \varphi_2(t\delta)$ for some positive t; and (3) if $n = 1$, transitivity is always satisfied. We note that for $n = 2$, if one of the dimension difference functions is nonlinear and the other is linear then the above transitivity condition does not hold and preferences may be intransitive.

The extended additive difference model (probabilistic)

In characterizing choice as probabilistic, repeated presentations of (x,y) in a two-alternative, forced choice paradigm are assumed to be independent Bernoulli trials, i.e., binomial random variables. Let p(x,y) be the probability of choosing lottery x over lottery y, and suppose that no choice probability is 0 or 1. Let F be an increasing function.

The extended additive difference model is satisfied whenever equation (1) holds and:

$$p(x,y) = F(osd(x,y)) = F(\sum_{i=1}^{n} \varphi_i[u_i(x_i) - u_i(y_i)])$$

Tversky comments that this is a Fechnerian model, such that choice variability is due to the relationship between the overall subjective difference between x and y, osd(x,y) and p(x,y). He suggested that function F can be either the normal or the logistic function – here we specify the logistic function:

$$p(x,y) = F(osd(x,y)) = \exp(osd(x,y))/(1+ \exp(osd(x,y))).$$

A simplified additive difference model for Tversky's lottery paradigm

In the following we present a simplified model specifically for Tversky's lottery paradigm.

Let A = S x P be a set of two-dimensional lottery alternatives of the form (s_i,p_i) such that payoff s_i is won with probability, p_i, otherwise nothing is won, i = 1, … 5. The set of payoffs, s_1, …., s_5 are decreasing in equal intervals, d_s, and the set of probabilities, p_1, …. , p_5 are increasing in equal intervals, d_p. We also denote the five lotteries of set A as a = (s_1,p_1), b = (s_2,p_2), c = (s_3,p_3), d = (s_4,p_4), e = (s_5,p_5). The specific values of Tversky's original lottery set are shown in Table 1.

We simplify the additive difference model as follows. For lottery set A = S x P we denote the scales u_1 and u_2 as u_s and u_p, the corresponding subjective values on dimensions S and P as $u_s(s)$ and $u_p(p)$ respectively, and the subjective dimension difference functions, $\varphi_1(\delta_1)$ and $\varphi_2(\delta_2)$ as $\varphi_s(\delta_s)$ and $\varphi_p(\delta_p)$. Next, we let $u_s(s) = s_i/d_s$ and $u_p(p) = p_i/d_p$, the difference in subjective values on dimension P be $\delta_p = (p_i – p_j)/d_p$, and the difference in subjective values on dimension S be $\delta_s = (s_i – s_j)/d_s$. Note that this standardizes the difference in subjective dimension values, δ_p and δ_s, to a common scale which is the objective standard difference level between lotteries of the set, d_c.

For lottery pairs in Tversky' paradigm, d_c takes the values 1, 2, 3, 4, as difference level increases. Difference functions are applied to the P and S dimensions, denoted $\varphi_p(\delta_p)$, and $\varphi_s(\delta_s)$. These can be viewed as the advantage or disadvantage of lottery x over y on the P and S dimensions respectively. Let $\varphi_p(-\delta_p) = -\varphi_p(\delta_p)$ and $\varphi_s(-\delta_s) = -\varphi_s(\delta_s)$. In the following we substitute the standardized objective difference, d_c, for δ_p and δ_s, since by our simplification, $\delta_p = \delta_s = d_c$. This means that the subjective dimension difference functions, φ_p and φ_s, and the overall subjective difference, osd, are all functions of d_c.

A preference structure in Tversky's lottery paradigm satisfies the simplified additive difference model if there exist continuous difference functions, $\varphi_p(d_c)$ and $\varphi_s(d_c)$ such that:

(2) $x \succcurlyeq y$ if and only if $\varphi_s(d_c) - \varphi_p(d_c) \geq 0$

where x has a lower P and higher S value than y in Tversky's lottery set, and d_c is the objective standard difference between x and y.

If one of the dimension difference functions is nonlinear and the other is linear then Tversky's transitivity condition described above is not met and preferences may be intransitive. This is illustrated in Figure 1, where subjective differences for the S dimension are linear, $\varphi_s(d_c) = d_c$, and the subjective differences on the P dimension are nonlinear, $\varphi_p(d_c) = 0.4d_c^2$. In this figure it can be seen that the overall subjective difference, $osd(d_c)$ is positive for $d_c = 1$ and $d_c = 2$ since $\varphi_s(d_c) > \varphi_p(d_c)$; in these cases, therefore, $x \succcurlyeq y$. However, for $d_c = 3$ and $d_c = 4$, since $\varphi_s(d_c) < \varphi_p(d_c)$, preference switches to $y \succcurlyeq x$. In this case, then, the simplified additive difference model predicts intransitivity of preferences.

To model the simplified additive difference model on Tversky's lottery set there are two approaches that could be adopted. First, it might be seen as more natural to specify and estimate parameters for the dimension difference functions, $\varphi_s(d_c)$ and $\varphi_p(d_c)$. However, we adopt the second, more elegant and parsimonious, approach and directly model the overall difference function, osd(d_c), as explained below.

------ Figure 1 about here ------

The two-parameter, simplified additive difference (SAD) model

Dimension difference functions that predict intransitive preferences in the manner illustrated in Figure 1 will yield overall subjective difference functions, osd(d_c), that are either monotone increasing or monotone decreasing, and change from positive to negative, or vice versa, in the range $1 < d_c < 4$. A curvilinear function of this type would predict intransitive preferences, including, as Tversky (1969) has observed, a step function consistent with the LS heuristic. However, a linear function is a more parsimonious function with these properties. Furthermore, it predicts intransitive preferences in line with those from a nonlinear function if both functions predict preference switches at the same difference level. Thus, we define the (algebraic) two-parameter, simplified additive difference (SAD) model for Tversky's lottery paradigm by the following linear function:

(3) $\quad osd(d_c) = a_0 + a_1 d_c$

where d_c is the objective dimension difference level between lotteries x and y described earlier.

The linear overall subjective difference function models the difference between subjective dimension difference functions that may be either nonlinear, or linear but not related by the equation $\varphi_s(d_c) = \varphi_p(td_c)$.

As indicated above, an important property of the model is that it enables us to predict when preferences will be transitive or intransitive. This is illustrated in figure 2, which shows the graphs of four specific two-parameter SAD models. In two of them the linear function crosses the horizontal axis, i.e. at $d_c = -a_0/a_1$, between $d_c = 1$ and $d_c = 4$, thus predicting that preference will switch from the better S to the better P in the range $1 < d_c < 4$, or vice versa. That is, if $1 \leq -a_0/a_1 \leq 4$ preferences will be intransitive; for example, if $-a_0/a_1 = 1.5$, $osd(d_c) > 0$ for pairs (a,b) and (b,c), but $osd(d_c) < 0$ for pair (a,c) in the lotteries of Tversky's paradigm. For the other two models shown in the figure, the value of $-a_0/a_1$ is not in this range and the model predicts that preferences will be transitive, since $osd(d_c)$ is either always positive, or always negative, in the range $1 < d_c < 4$.

----- Figure 2 about here ----

As with the extended additive difference model, the extended (probabilistic) two-parameter SAD model is satisfied whenever equations (2) and (3) hold and

$$p(x,y) = F(osd(d_c)) = \exp(osd(d_c))/(1+ \exp(osd(d_c)))$$

where d_c is the objective difference level for lottery pair (x,y).

The extended model predicts that WST will be violated when $1 \leq -a_0/a_1 \leq 4$, and conversely, that WST will hold when $-a_0/a_1$ is outside that range. For example, in the case described above, where $-a_0/a_1 = 1.5$: p(a,b) > .5, p(a,c) >.5 but p(a,c) < .5, which violates WST[3].

Other models against which the extended two-parameter SAD model can be tested

We compare the goodness of fit of the extended two-parameter SAD model on choices in Tversky's paradigm to that of four others in a family of nested probabilistic models.

The ten-parameter baseline, model (M_0). Following Tversky (1969) and others, we assume that probabilistic models with less than ten parameters are nested within a baseline model that Tversky referred to as the nonrestrictive model, M_0. This places no restrictions on the ten

[3] The SAD model also predicts when the TI condition is satisfied or violated, although the boundary conditions for this are not straightforward.

binary choice probabilities of Tversky's lottery paradigm, and consequently has ten parameters, one for each pair of lotteries in the set. Cavagnaro and Davis-Stober (2014) refer to this as the encompassing, or baseline mode. If M_0 is not a significantly better fit than a model with fewer parameters, such as the two-parameter SAD model, we conclude that the latter has a good fit to the data.

The four-parameter cubic model. This is another baseline model that provides us with a test of departures from the linearity assumed by the two-parameter SAD model:

(4) $osd(d_c) = a_0 + a_1 d_c + a_2 d_c^2 + a_3 d_c^3$

If the extended cubic model is a good fit and significantly better than the SAD model, then we should conclude that the data does not support the linear model. However, the further interpretation of a well-fitting cubic model depends on its specific parameter values. Parameter values that approximate a step function, are monotonic increasing or decreasing and change from positive to negative, or vice versa, in the range $d_c = 1$ to $d_c = 4$ can be interpreted as being consistent with the LS heuristic predicting intransitive preferences in the sense of violating WST.

Others parameter combinations, however, can indicate that the relationship between $osd(d_c)$ and d_c is nonmonotonic, and that the dimensional models considered here should be rejected.

The one-parameter constant choice probability (CCP) model. In this model the slope of the linear relationship of the SAD model is zero:

(5) $osd(d_c) = a_0$.

The extended CCP model predicts that across the ten binary pairs of the lottery set the binary choice probability will be constant and therefore, WST will be satisfied. Since choice probability is not sensitive to differences between P or S values under this model, it is a decision heuristic. Figure 3 illustrates how the model can be interpreted for different values of a_0: a value $a_0 > 0$ reflects a preference for the higher S, whereas $a_0 < 0$ reflects a preference for a higher P; $a_0 = 0$ defines indifference, corresponding to random choice. Additionally, a value of $a_0 \leq -1.39$ represents a stochastic 'take the best P' heuristic, and $a_0 \geq 1.39$ represents a stochastic 'take the best S' heuristic (Gigerenzer & Goldstein, 1999). Since the CCP model predicts transitive preferences, observed intransitive choice cycles are explained by the inherent stochastic nature of choice, with choice probability being related to the strength of preference for S over P or vice versa.

------ Figure 3 about here -----

The zero-parameter random choice (RC) model. As mentioned above, if $a_0 = 0$ in equation (5) the extended model predicts indifference across all pairs of the lottery set and, therefore, random choice. Although it predicts that WST will hold, this is a case of the transitivity of indifference.

The focus of our analysis is the two-parameter SAD model, in particular the number of cases where it is a significantly better fit to the data than the CCP model, is a good fit to the data, WST is predicted to be violated, and neither the cubic model nor M_0 are significantly better. Alternative patterns of choice proportion exhibiting violations of WST predicted by the SAD model are illustrated in the bottom panels of Table 1. This shows that the main difference with the predictions of the LS heuristic (upper panels) is a more gradual change in choice proportions as dimension differences increase.

Tversky's (1969) lottery experiment and six replications

Tversky (1969) invited 18 students to participate in his lottery study. In the first session they chose between adjacent lotteries from the set shown in Table 1, i.e. pairs (a,b), (b,c), (c,d) and (d,e) as well as the extreme pair, (a,e). Each of these was presented three times with filler pairs interspersed, and afterwards participants played the gamble chosen on a randomly selected trial for real. Eight participants then returned for the main experiment, all those who had chosen the higher S on the majority of adjacent pairs at least twice, but the higher P on the extreme pair at least twice. Regenwetter et al. (2011) describe this selection of potentially intransitive participants as “cherry picking”. However, we find that the sample selected from, and the method of selection, are clear and transparent, and consistent with the aims of the study. In the main experiment participants chose from each of the ten pairs from the lottery set 20 times across five sessions, one week apart. As before, filler choices were interspersed and at the end of a session participants played one chosen lottery for real. Tversky reported that likelihood ratio tests showed that the choice patterns of five participants significantly violated WST at the 5% level of significance or lower, while only one participants’ pattern violated the LS heuristic. The choice patterns of two of Tversky’s participants are shown in Table 2.

------- Table 2 about here -----

Regenwetter et al. (2011) challenged Tversky’s conclusions on several grounds, one of which cited Iverson and Falmagne’s (1985) reanalysis that identified a flaw in Tversky’s likelihood ratio test of WST. Iverson and Falmagne derived a valid alternative test which found that only one of Tversky’s participants significantly violated WST. Regenwetter et al. themselves came to a similar conclusion: the choice pattern of only one participant was significantly outside the TI condition at the $p < .05$ level, with one other being borderline at $p = .05$. Thus, their transitive preferences model, the linear order mixture model, could be rejected for only two of

Tversky's 18 participants at most. However, using Bayes factor analysis, Cavagnaro & Davis-Stober (2014) found strong evidence against transitive preferences for three participants, with evidence supporting transitive preference being inconclusive for three others (their analysis tested both the mixture model and WST). In summary, while Tversky's original analysis found that 5/18 participants exhibited significant intransitive preferences, others have concluded that the number is actually fewer, ranging from one to three out of 18 participants.

We reanalyzed Tversky's (1969) lottery experiment and six replications. The three earlier replications, Montgomery (1977), Ranyard (1977) and Tsai and Böckenholt (2006), were reviewed by Regenwetter et al. (2011 and by Cavagnaro & Davis-Stober, (2014), who also conducted their own replications. Kalenscher et al.'s (2010) experiment was reanalyzed by Brown, Davis-Stober & Regenwetter (2015). Five of the studies adopted the five lotteries that Tversky devised, adjusted for currency and inflation. The exception was Tsai and Böckenholt's that only presented the first four lotteries. This was also the only study that did not have a real consequence of a lottery choice, and it was the one that presented each lottery pair the most times, 120, rather than the range 8 to 20 of the other studies. Another important difference among the studies is that Montgomery (1977 and Ranyard (1977) only fully analyzed a subset of participants, whereas the other studies fully analyzed the data from the whole sample. Further details of the six studies, and previous reviews of them, are presented in appendix 1.

Method

Our reanalysis of the seven data sets described above has two phases. In the first, exploratory phase, we conduct a descriptive analysis of individual choice proportions. An important aspect of this is to identify participants whose choices had very little variability. In these cases choice is close to deterministic and decision strategies can be clearly identified without recourse to probabilistic analysis, which in any case would be inappropriate. For the remaining participants we move to the second, probabilistic model testing phase.

Exploratory analysis

Previous studies of choice in Tversky's (1969) lottery paradigm did not include a detailed descriptive analysis of individual binary choice proportions. Tversky himself merely presented the matrix of choice proportions from the lottery set for eight selected participants and identified those that were inconsistent with either WST or an LS heuristic. Montgomery (1977) adopted the same descriptive approach, while Ranyard (1977) only presented the choice frequencies of selected participants. Regenwetter et al. (2011) also only presented choice frequencies, although these were in the supplementary material available from the authors rather than in the main paper. For the more recent studies we had to calculate the choice proportions from data in an appendix (Tsai and Böckenholt, 2006) or request the information from the authors (Cavagnaro & Davis-Stober, 2014; Kalenscher et al., 2010). [4]

Our descriptive measures are conventional ones that were not inspected closely in the original studies: means and confidence intervals for the ten choice proportions of the set, and a regression coefficient measuring the linear relationship between dimension difference and choice proportion. For each participant we first we calculated statistics relevant to the CCP model: (1) the mean choice proportion (cp) across the set; (2) the 95% confidence interval (CI) for cp under the CCP model, identified from the binomial distribution for each mean cp; and (3) those cps lying outside the CI, noting the number of outliers, their position in the set, and their

[4] All data are available online on the Open Science Framework (OSF) website: https://osf.io/y8fxg/

probabilities of occurrence under the CCP model. Second, we calculated a descriptive statistic relevant to the SAD model, which predicts that across the ten choice proportions of the set there will be a relationship between the level of dimension difference, d_c, and cp. Specifically, we calculated the unstandardized regression weights (B) for the linear regression predicting cp from d_c. Finally, we noted whether violations of WST occurred for cycles of three, four or five binary choice pairs occurred, as predicted by LS heuristics with different thresholds.

As well as summarizing the basic characteristics of individual choice data, the exploratory analysis identified those cases with minimal choice variability that could be classified on the basis of an algebraic dimensional model with very low error rate. Specifically, those where the mean choice proportion is very low or very high, with many being zero or many being one, were classified as ‘take the best S’ or ‘take the best P’. In these cases probabilistic model fitting is neither necessary nor appropriate.

Probabilistic models: parameter estimation, tests and categorization

Categorization of individual choice behavior

The main conditions defining different categories of choice behavior are described below. As discussed earlier, under the two-parameter SAD model, preferences may satisfy or violate WST depending on the specific values of the model’s parameters. Also as mentioned earlier, the one-parameter CCP model is a transitive case of the SAD model when $a_1 = 0$. The categories are defined in terms of the parameter values of the overall subjective difference function, $osd(d_c) = a_0 + a_1 d_c$, where d_c, the objective difference between lotteries in Tversky’s task, varies from 1 to 4. Three main cases can be distinguished, with four variants of the third case, including the baseline random choice model (codes for each case in parentheses). For the first two categories the extend two-parameter SAD model is a good fit to the data. The model predicts that WST will be violated when $1 \leq -a_0/a_1 \leq 4$, and conversely, WST will hold when $-a_0/a_1$ is outside that range.

1. Intransitive preferences.

   If $1 < -a_0/a_1 < 4$, then preferences will violate WST (SAD INT).

2. Transitive preferences with $a_1 \neq 0$.

   If $-a_0/a_1$ is outside the above range, preferences will be transitive, and satisfy WST (SAD WST).

For the third category, and its subcategories, the extended one-parameter CCP model is a good fit and models with more parameters are not significantly better. Note that the values $a_0 = \pm 1.39$ are those predicting choice proportions of 20 or 80 percent.

3. Transitive preferences with $a_1 = 0$

   a. Constant choice probability other, if $-1.39 < a_0 < 1.39$ (CCP(O))

   b. Take the best P, if $a_0 \leq -1.39$ (CCP(P))

   c. Take the best S, if $1.39 \leq a_0$ (CCP(S))

   d. Random choice, if $a_0 = 0$. (RC)

Finally, we identify two categories of choice behavior when the cubic model is a good fit and significantly better than the SAD model. As stated earlier, this indicates significant nonlinearity of overall subjective differences such that the SAD model is rejected. Classification is based on visual inspection of choice proportions.

4. Nonlinear overall subjective differences

   a. Intransitive preferences, nonlinearity approximating a step function consistent with the LS heuristic that predicts violation of WST (LS INT)

   b. Transitive preferences, nonlinearity which is nonmonotonic (NM).

Model fitting

The more powerful maximum likelihood (ML) method is preferred because it uses all the information in the data by estimating parameters from individual choice trial data, rather than a least squares method fitting choice proportions without reference to the number of replications (see Kalenscher et al., 2010). Furthermore, with the ML method, likelihood ratio (LR) tests can be used to test the goodness of fit of the different models described earlier.

Maximum likelihood estimation (MLE). For the cases that were designated probabilistic after exploratory analysis, we applied the ML method to estimate model parameters. As mentioned earlier, we assume that repeated presentations of any two lotteries, (x,y), in a two-alternative, forced choice paradigm are independent Bernoulli trials. Each choice trial is an independent, binomial random variable, where p(x,y) is the probability of choosing lottery x over lottery y. We assume that no choice probability is 0 or 1. If each lottery pair is presented r times, the likelihood of the observed choice frequency, i.e. the number of times x is chosen over y out of r presentations, given a predicted p(x,y), is determined by the Binomial distribution. Under the further assumption that the binary choice probabilities are independent, the overall likelihood is given by the product of the likelihoods of the ten choice proportions. In our MLE algorithm to estimate parameter values we let choice proportions of zero or one take the values .01 and .99 respectively, in line with best MLE practice (Myung, 2003).

We use the statistic -2ln(likelihood), denoted -2LL, to assess the goodness of fit of the resulting models. Under the standard ML approach, this can be assumed to approximate a chi-square statistic with degrees of freedom equal to the number of parameters of the model. We also applied the ML method to the one-parameter CCP model by finding the MLE of parameter a_0 when $a_1 = 0$.

Model testing and classification of individual choice data. We applied LR tests to determine the best fitting model of the five nested models described earlier (number of parameters in parentheses): M_0 (ten), Cubic (four), SAD (two), CCP (one) and RC (none). In these tests, the difference in -2LL between two models is assumed to approximate a chi-square distribution with degrees of freedom equal to the difference in the number of parameters of the models. In comparisons with M_0, if this statistic is less than the critical value for $p = .05$, we conclude that the model has a good fit to the data. In comparisons between the other models, if the chi-square statistic is significant at $p < .05$, we conclude that the model with the higher number of parameters has a significantly better fit. Otherwise, we conclude that the additional parameters do not significantly improved model fit. The aim is to identify the model with the lowest number of parameters that is a good fit to the data, not improved upon significantly by a model with more parameters. By the principle of parsimony, we conclude that out of all models with a good fit, the one with the lowest number of parameters is the best fit. For example, if no model has a significantly better fit than the RC model we conclude that the RC model is a good, and the best fit. We then test the CCP model against models with more parameters, then the SAD model, and finally the cubic model, stopping if we identify a model that is a good, and the best fit. The outcomes of these tests, together with the MLEs of parameters, determine our categorization of choice data as described earlier. The tests also reveal cases where none of the models tested against M_0 have a good fit. Additionally, where the Cubic model is a good fit and the best fitting model, it is interpreted in a follow-up exploratory analysis, specifically, a visual inspection of the scatterplot of the log-odds of choice proportions, to determine whether it should be classified as a case of intransitive preferences, i.e., consistent with the LS heuristic and violating WST (LS INT), or as transitive preferences but with a nonmonotonic relationship between choice proportion and level of dimension difference (NM).

Results

Tversky (1969)

Our reanalysis is summarized in Table 3, with the descriptive statistics of the exploratory analysis to the left and model fitting information to the right. The left of the table shows that for the first six participants there are several outlier cps, a strong linear relationship between cp and d_c, and observed violations of WST, suggestive of intransitive preferences consistent with the SAD or LS models. On the other hand, the statistics for participant seven are more suggestive of the CCP model, since there are no outlier cps, a very weak relation between cp and d_c, and only one observed violation of WST. Finally, for participant eight the statistics suggest a transitive pattern with no observed violations of WST and a strong relationship between cp and d_c consistent with the SAD model (see Table 2).

These observations are broadly confirmed by the ML estimation and goodness of fit results shown to the right of the table. For five participants (1 to 4 and 6) the difference between the -2LL values for M_0 and the SAD model, and between the cubic and SAD model, were not significant, but those between SAD and CCP, and between SAD and RC, were significant. In these cases, then, the SAD model is a good, and the best fit, with values of a_0 and a_1 predicting systematic violations of WST (i.e., $1 < -a_0/a_1 < 4$). For participant 5, however, while the SAD is a good fit and significantly better than the CCP model, the cubic model is significantly better than the SAD model, indicating nonlinearity. Inspection of choice proportions shows that this is due to a nonmonotonic relationship such that the SAD model is not supported (classified as NM).

For participant seven, the CCP model has a good fit and is the most parsimonious, since the two-parameter SAD model is not significantly better. Finally, for participant eight the SAD model is a good, and the best, fit with parameter values consistent with transitive preferences. We conclude, therefore, that the SAD model predicting violations of WST is a good, and the best fit for 5/18 of Tversky's participants.

---- Table 3 about here ------

Early replications

The results presented in Table 4 show that for Montgomery (1977) and Ranyard (1977) the SAD model was a good, and the best fit for six of the seven participants fully reported, with parameter estimates for these six predicting the observed violations of WST. The exception was Montgomery's participant five. In this case, while the SAD model was a significantly better fit than the CCP model, inspection of the data reveals that its relatively poor fit in comparison to M_0 was due to variations in choice proportion within some levels of dimension difference.

Turning to Tsai and Böckenholt (2006), for two participants the random choice (RC) model was a good, and the best fit. However, in three other cases, the SAD model was a good, and the best fit with parameter estimates predicting the observed violations of WST (see Table 6, top panels for examples). Comparing the first two studies with the later one, the rate of change of choice proportion as dimension difference changes is rather lower in Tsai and Böckenholt's cases, around 10 percent compared to 15-30 percent in the earlier studies. Nevertheless, overall for about a third of participants in these three studies the SAD model predicting observed violations of WST was a good, and the best fit.

----- Table 4 about here ------

Regenwetter et al. (2011)

We present the reanalysis of the Cash 1 data, which was a replication of Tversky's original lotteries adjusted for inflation. Our exploratory analysis (Table 4, columns 2 to 6) found that six participants (3, 5, 8, 10, 11 and 14) could be classified as being consistent with take the best P or take the best S heuristics on the criteria described earlier. For the remaining 12 participants, probabilistic model testing using MLE and LR tests was carried out (Table 4, from column seven). This identified two further cases where the CCP model was a good, and the best fit: participant 7 was classified as take the best P; and participant 9 was classified as CCP(O). These were so classified because: (1) the differences between -2LL for this model and either M_0, the cubic model or the SAD model were not significant ($p > .05$); and (2) the differences between -2LL for this model and RC were significant. In addition, we identified seven cases where the SAD model was a good fit, with neither M_0 nor the cubic model being significantly better ($p > .05$), but where the SAD model was significantly better than the CCP or RC models ($p < .05$). Four of these predict the observed violations of WST (see Table 6, bottom panels for examples) while for the other three transitive preferences are predicted. Finally, there were three cases for which none of the models tested was a good fit, due to substantial variability in choice proportions for some dimension differences. The most important finding, however, is that 4/18 cases exhibited violations of WST predicted by the well-fitting SAD model.

----- Tables 5 and 6 about here -----

Cavagnaro and Davis-Stober (2014)

This study presented two sets of lotteries to participants with or without time pressure. We reanalyzed Set 1 data in the without time pressure condition since this replicates Tversky's (1969) experiment. Our exploratory analysis, presented in Table 7 columns two to six, shows that eight participants were classified as consistent with 'take the best' heuristics. For the remaining 21 cases probabilistic model testing is summarized in the table from column seven. This identified two participants classified as CCP(O) and four as SAD (WST), giving a total of 14 (48 percent) for which transitive variants of the models were the best fit. In a further four cases, none of the models tested was a good fit, due to variability of choice proportions for some dimension differences. In another case the cubic model was a good, and the best fit, because of nonlinearity of choice proportions across dimension difference (classified as NM). However, the most important finding is that for ten participants (34 percent) the SAD model was a good, and the best fit with parameter values predicting the observed violations of WST.

---- Table 7 about here ------

Kalenscher et al. (2010)

As indicated in Table 8, exploratory analysis identified four cases not suitable for probabilistic modelling because of very high or low mean choice proportions. These were classified as consistent with take the best P or S heuristics. Of the remaining 26, the cubic model was a good, and the best, fit for a relatively high number: the seven participants asterisked in the M_0 column of the table. Inspection of the configuration of choice proportions in these cases showed that violations of WST consistent with the LS heuristic were observed in three (classified as LS INT) and transitive preferences with a nonmonotonic relationship between choice proportion and dimension difference in four (classified as NM). These latter cases do not fit the dimensional models under investigation because of significant nonmonotonicity. In addition, poor fit of dimensional models was identified in six other cases where M_0 was the best fit. In the remaining 13 cases the SAD model was a good, and the best fit, with two being transitive (classified as SAD WST) and 11 predicting observed violations of WST (SAD INT). Overall, then, for 14/30 (47 percent) of participants a dimensional model predicting violations of WST was a good, and the best fit to the data, either the SAD model or the cubic model consistent with the LS heuristic.

----Table 8 about here -----

Overview of the reanalysis

An overview of our reanalysis is presented in Table 9. The first column shows the percent of participants in each study for whom the two-parameter SAD model was a good, and the best fit to the choice data, with parameter values predicting the observed violations of WST. In the case of Kalenscher et al.'s study, this includes three cases where the cubic function approximating the LS heuristic was a good, and the best fit, and with observed violations of WST consistent with this. We find a remarkable consistency across the seven studies reviewed with respect to the proportion of participants displaying violations of WST predicted by well-fitting dimensional models; one third of all participants across all studies.

The second column of Table 9 shows that for a further eight percent of participants overall, the two-parameter SAD model was also a good, and the best fit, although in these cases observed transitive preferences were predicted by the model. The third column shows that for about one in four participants from studies where full data sets were available the CCP model was a good, and the best fit, with observed transitive preferences predicted by the model. Most of these were classified as consistent with take the best S or P heuristics. In the fourth column it can be seen that random choice was a good, and the best fit for only two participants, both from Tsai and Bockenholt's (2006) study. Finally, despite the overall success of the dimensional models in predicting the data, the fifth column of Table 9 shows that for about fifteen percent of participants either M_0, or the cubic model with observed nonmonotonicity, had significantly better fit than the other models tested. This latter point represents the main limitation of an otherwise successful test of dimensional models.

----- Table 9 about here ----

Comparison with previous analyses

Previous analyses of Tversky's (1969) lottery study focused on establishing whether the individual choice data significantly violated WST and/or the TI condition. While Tversky concluded that 5/18 of his participants significantly violated WST, others subsequently disputed this. Notably, Regenwetter et al. (2011) concluded that only 1/18 significantly violated the above transitivity conditions, interpreting this violation as a type I error from a transitive preference model. However, Cavagnaro and Davis-Stober's (2014) Bayes factor analysis found that in 3/18 (17%) cases there was strong evidence against transitivity, in that both transitivity conditions were violated, with a further three cases being inconclusive. Our own reanalysis was consistent with this, since for five participants the two-parameter SAD model was a good, and the best fit, with parameter values predicting the observed violations of WST. Taking both of these analyses into account, we conclude that for three participants the additive difference model predicting observed violations of WST is supported conclusively, and transitive models (e. g., expected utility or cumulative prospect theory) are rejected. In addition, two other cases strongly support the SAD model predicting observed violations of WST, although these are inconclusive since a transitive model also fits the data well.

Turning to the three earlier replications, each claimed to have extended Tversky's (1969) findings in different ways. Ranyard (1977) identified cases where violations of WST were still evident when the probability information format was numerical rather than graphical, thereby eliminating perceptual issues. Montgomery (1977) also identified several cases violating WST, and presented a rigorous analysis of think aloud evidence of underlying dimensional processing. Both Regenwetter et al.'s. (2011) and Cavagnaro and Davis-Stober's (2014) reanalyses corroborated Ranyard's and Montgomery's findings. In particular, the latter found strong evidence against both WST and the mixture models in six cases across the two studies, while we found that the SAD model predicting violations of WST was a good fit in five cases. In these five cases, therefore, the explanation of intransitive preferences provided by the SAD model is supported unambiguously.

With respect to the third earlier study, while Tsai and Böckenholt (2006) did not test transitivity directly, they did present evidence from correlations between choices across gamble pairs that supported dependent response models, of which Tversky's additive difference model is one. Neither Regenwetter et al. nor Cavagnaro and Davis-Stober discussed these findings in their reviews, instead focusing on their own analyses, which found strong evidence supporting transitive preference models. In contrast, our reanalysis identified predictable violations of WST in three of Tsai and Böckenholt's five participants (not previously reported), and model testing showed that the SAD model predicting this was a good, and the best fit to the data. It should be noted, however, that this was the only replication that did not motivate participants by paying out on a randomly selected chosen lottery, which may explain why two participants apparently responded at chance level.

Let us consider further the discrepancy between these very different conclusions from different analyses of Tsai and Böckenholt's (2006) data. Regenwetter at al.'s conclusion, confirmed by Cavagnaro and Davis-Stober, was that all five participants' choice patterns are a good fit to the transitive mixture model, and do not significantly violate either the TI or WST conditions. We do not dispute this, but argue that our reanalysis gives further insights. First, as mentioned above we found that the random choice model (not previously tested on this data) was a good fit for the choices of two participants. While this is a transitive model, it is perhaps better characterized as transitivity of indifference. Second, as stated above, we observed violations of WST, with parameter estimates of the well-fitting SAD model predicting this. The data is inconclusive in these cases, since it is a good fit to both transitive and intransitive preference models. Nevertheless, taking Tsai and Böckenholt's own results into account, we suggest that the balance of evidence swings in favor of the dimensional processing model predicting violations of WST.

Turning to Regenwetter et al.'s (2011) study, we found that the SAD model predicting intransitive preference was a good fit for 4/18 of Cash 1 choice patterns (see Table 6 for individual data), a proportion of the sample not dissimilar to Tversky's study, in which at least 5/18 participants had a good fit to the model in the same way. From our perspective, then, this warrants a conclusion of replication rather than a failure to replicate. On average the choices of these four of Regenwetter et al.'s participants were quite sensitive to changes in dimension difference (as measured by the slope parameters of the SAD model, and the unstandardized regression coefficients, B, which ranged from -.11 to -.19, see table 5). Note that for B = -.15, choice proportion changes on average by 15 percent for each change in dimension difference. However, Cavagnaro and Davis-Stober's analysis found strong evidence against transitivity for only one of these cases (participant four) with the three others either being inconclusive or having strong evidence supporting transitivity. Taking all the analyses together, it seems that one of the four cases we identified gives unambiguously strong evidence supporting the SAD model predicting observed violations of WST, while the other three are inconclusive, since they can be explained by both transitive and intransitive models. In other respects our analysis confirms that most participants exhibited transitive preferences. However, a new insight from our analysis is that dimensional processing models predicting transitivity were a good fit for over half the sample (11/18, 61 percent), and most of these (8/18, or 44 percent) were consistent with the application of simple, take the best heuristics in which only one dimension is processed. Finally, inspection of the 3/18 cases for which the SAD model was not a good fit revealed that this was because of substantial variability of choice proportions for some particular dimension differences. Notwithstanding, we found that dimensional processing models successfully accounted for 83 percent of the sample.

In Cavagnaro and Davis-Stober's (2014) study we again find replication of Tversky's (1969) findings, since the two-parameter SAD model predicting violations of WST was a good, and the best fit for about one third of participants (10/29), again similar to the proportion of the sample in Tversky's study. However, four of these are inconclusive since Cavagnaro and Davis-Stober found strong evidence against transitivity in only 6/29 (21 percent). As in Regenwetter et al.'s (2011) study, we found here that for a substantial proportion of transitive participants, choice data were consistent with the use of simple, one-dimensional heuristics. Finally, a little disappointingly, the dimensional models did not have a good fit for 5/29 (17 percent) of cases.

Turning to Kalenscher et al.'s (2010) study, although for about one third of participants dimensional models were not a good fit to the data, again we find replication of Tversky's (1969) findings. Specifically, for 47 percent of participants (14/30) the SAD model or the LS heuristic were a good, and the best fit with observed violations of WST being predicted by these models. However, while support for Tversky's models was unambiguous in seven cases, the other four should be seen as inconclusive, since Brown et al. (2015) found that only 21 percent (7/30) of cases strongly supported intransitivity. Thus, both transitive and intransitive models fit the data of these four participants. It would be interesting to reinterpret Kalenscher et al.'s evidence concerning neural correlates in the light of our classification, bearing in mind that different categories imply the processing of more or less information.

Overall, then, we have shown how the six studies reviewed replicated Tversky's (1969) initial findings, with about 30 percent exhibiting violations of WST predicted by the SAD model. Although in a third of these cases the evidence is inconclusive, since transitive models also fit the data well, we conclude that in other two thirds there is unambiguous strong support for Tversky's (1969) models predicting observed violations of WST.

General discussion

Previous reviews of Tversky's (1969) lottery study and its replications have viewed them through the lens of transitive models such as the weak utility or random utility models. From this perspective, Regenwetter et al. (2010, 2011) interpreted observed intransitive choice cycles as type I errors. While Cavagnaro and Davis-Stober (2014) did not concur, they did conclude that choice behavior in the lottery task was broadly transitive, such that evaluations of decision alternatives are basically stable. In our own review we have challenged this characterization, as Tversky did, by looking at the evidence through the lens of dimensional processing models. This assumes that the evaluation of decision alternatives can be relative, and can change as the set of alternatives under consideration changes. From the dimensional processing perspective, we showed how evaluation can be relative, and choices can violate WST, if decision makers process dimension differences on both payoff and probability dimensions. However, preferences will be transitive and stable if only one of these dimensions is processed. With respect to the latter point, we showed that a substantial proportion of transitive choices were consistent with simple, one-dimensional heuristics. All this was supported by our exploratory analysis, inspired by Tukey's (1980) approach, and formalized in our specification of the SAD model and its variants, which accounted for about 85 percent of the data.

We would like to make clear some caveats. First, the SAD model is a testable simplification of Tversky's (1969) extended additive difference model specifically designed for his lottery task. Therefore, to test the model in a wider context alternative overall difference functions would need to be developed, preferably modelling each subjective dimension difference function. Second, our treatment of stochastic transitivity of choice and preference is mainly based on WST rather that the TI condition related to random utility models. The insightful critique of WST by Regenwetter et al. (e.g., 2010) is important, but we believe that retaining WST as a criterion of stochastic transitivity in evaluating Tversky's extended additive difference model for intra-individual preference is justified for the following reasons. Their first criticism, that WST is confounded with variability of preference (p.6), is justified for data aggregated across individuals, since preferences do vary across individuals. However, the SAD model assumes that an individual's underlying pairwise preferences are not variable, with stochastic binary choices predicted by a logistic function of strength of preference. That is, if context is fixed, as in Tversky's lottery paradigm, the model assumes that individual preferences between pairs of gambles are invariable but stochastic (see Tversky, 1969). Their second point, that WST does not model transitivity in isolation from other axioms of preference is well-argued and demonstrated, and we accept that violations of WST could be due to violations of the other axioms of a weak order, reflexivity and completeness. However, this only means that violations of WST might be due to violations of these other axioms. Turning to their third point, that WST only allows linear preference orders…. *and … neglects … transitive relations that are …not* weak orders, this argues that WST is only a partial test of stochastic transitivity. Their fourth criticism is that WST treats probability as a binary category rather than an absolute scale. This is true, but the extended additive difference model (and therefore the SAD model) models choice probability on a ratio scale, and when WST is treated as a testable consequence of this fully probabilistic model its binary status is not relevant. For the above reasons we conclude that despite some limitations WST remains an important criterion of stochastic transitivity. Notwithstanding the above, to give a more nuanced comparison of models than we have so far been able to achieve, future research could examine the extent to which the SAD model predicts observed violations of the TI condition. Our final caveat is to note that our comparison of the present analysis with previous ones is only definitive with respect to the majority of cases that are interpreted as either unambiguously transitive or unambiguously intransitive.

We conclude with some suggestions for further research. First, it is important to investigate why some participants' choices did not fit dimensional models because of variability of choice proportions for a given level of dimension difference. It is possible that this is due to the use of hybrid dimensional and absolute, or holistic, evaluation of payoffs, probabilities or lotteries. This could be explored further with process tracing methods. Second, as discussed earlier, it would be instructive to revisit Kalenscher et al.'s (2010) work on neural correlates of lottery decision making in the light of our findings. Third, in addition to the need for further research in laboratory settings, a natural next step would be to extend the present research to the real world to reveal when and why people sometimes are intransitive and sometimes are transitive in their choices in certain domains, such as financial investments, consumer choice and social preferences. Fourth, our analysis of the four-parameter cubic model found that this approximation to a lexicographic semiorder was only in a few cases significantly better at predicting violaions of WST than the SAD model. A direct comparison of the SAD model with Davis-Stober's (2012) stochastic lexicographic models would be a better way to clarify the predictive power of the two models (Davis-Stober, Brown & Cavagnaro, 2015; Regenwetter, Dana, Davis-Stober & Guo, 2011). Related to this, it may also be useful to directly compare the SAD model with other developments of Tversky's additive difference model (e.g. Loomes, 2010; Rubinstein, 1988; Leland, 1994). Since these developments have been focused on deterministic representations (cf. Loomes, 2010, p. 1), an interesting task for future research would be to evaluate the predictive performance of the SAD model compared to stochastic versions of these more recent models. In this we are consonant with Loomes (2010, p. 14) who stated with regard to his PRAM model that "the incorporation of a stochastic element … could be profitably investigated in future research".

In conclusion, we have argued that an exploratory, descriptive analysis is essential for a rigorous investigation of the replicability of empirical findings. A second requirement is to compare any reanalysis from a new perspective with that from the original perspective. This is what we have done, and thereby provided clear evidence that Tversky's (1969) original findings are robust and replicable, and that his contribution to theory with respect to dimensional processing in decision making is valid and insightful.

References

Beach, L. R., & Mitchell, T. R. (1978). A contingency model for the selection of decision strategies. Academy of management review, 3, 439-449.

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

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: making choices without trade-offs. Psychological review, 113(2), 409-432.

Brown, N., Davis-Stober, C. P., & Regenwetter, M. (2015). Commentary: "Neural signatures of intransitive preferences". Frontiers in human neuroscience, 9.

Cavagnaro, D. R., & Davis-Stober, C. P. (2014). Transitive in our preferences, but transitive in different ways: An analysis of choice variability. Decision, 1, 102-122.

Davis-Stober, C. P. (2012). A lexicographic semiorder polytope and probabilistic representations of choice. Journal of Mathematical Psychology, 56, 86-94.

Davis-Stober, C. P., & Brown, N. (2011). A shift in strategy or" error"? Strategy classification over multiple stochastic specifications. Judgment and Decision Making, 6, 800-813.

Davis-Stober, C. P., Brown, N., & Cavagnaro, D. R. (2015). Individual differences in the algebraic structure of preference. Journal of Mathematical Psychology, 66, 70-82.

Edwards, W. (1954). The theory of decision making. Psychological Bulletin, 51, 380.

Gigerenzer, G., & Goldstein, D. G. (1999). Betting on one good reason: The take the best heuristic. In Simple heuristics that make us smart (pp. 75-95). Oxford University Press.

Glöckner, A., & Herbold, A. K. (2011). An eye-tracking study on information processing in risky decisions: Evidence for compensatory strategies based on automatic processes. Journal of Behavioral Decision Making, 24(1), 71-98.

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

Kahneman, D. (1979). Tversky A.(1979). Prospect theory: an analysis of decision under risk, 263-292.

Kalenscher, T., Tobler, P. N., Huijbers, W., Daselaar, S. M., & Pennartz, C. M. (2010). Neural signatures of intransitive preferences. Frontiers in Human Neuroscience, 4.

Leland, J. W. (1994). Generalized similarity judgments: An alternative explanation for choice anomalies. Journal of Risk and Uncertainty, 9,151-172.

Loomes, G. (2010). Modeling choice and valuation in decision experiments. Psychological Review, 117(3), 902-924.

Luce, R. D., & Suppes, P. (1965). Preference, Utility, and Subjective Utility. Handbook of Mathematical Psychology, III (pp. 249-409). New York: Wiley.

Müller-Trede, J., Sher, S., & McKenzie, C. R. (2015). Transitivity in context: A rational analysis of intransitive choice and context-sensitive preference. Decision, 2(4), 280-305.

Montgomery, H. (1977). A study of intransitive preferences using a think aloud procedure. In H. Jungermann & G. de Zeeuw (Eds.), Decision making and change in human affairs. Dordrecht: Reidel.

Montgomery, H., & Svenson, O. (1976). On decision rules and information processing strategies for choices among multiattribute alternatives. Scandinavian Journal of Psychology, 17, 283-291.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. Journal of mathematical Psychology, 47(1), 90-100.

Oliveira, I.F.D., Zehavi, S, & Davidov, O. (2018). Stochastic transitivity: Axioms and models. Journal of Mathematical Psychology,85, 25-35.

Payne, J. W. (1982). Contingent decision behavior. Psychological Bulletin, 92(2), 382.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14, 534.

Pachur, T., Hertwig, R., & Wolkewitz, R. (2014). The affect gap in risky choice: Affect-rich outcomes attenuate attention to probability information. Decision, 1(1), 64.-78.

Ranyard, R. H. (1976). An algorithm for maximum likelihood ranking and Slater's i from paired comparisons. British Journal of Mathematical and Statistical Psychology, 29, 242-248.

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

Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2010). Testing transitivity of preferences on two-alternative forced choice data. Frontiers in psychology, 1, 148.

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

Regenwetter, M., Dana, J., Davis-Stober, C. P., & Guo, Y. (2011). Parsimonious testing of transitive or intransitive preferences: Reply to Birnbaum (2011). Psychological Review, 118, 684-688.

Regenwetter, M., Davis-Stober, C. P., Lim, S. H., Guo, Y., Popova, A., Zwilling, C., Cha, Y.-C., & Messner, W. (2014). QTEST: Quantitative Testing of theories of binary choice. Decision, 1, 2-34

Rubinstein, A. (1988). Similarity and decision-making under risk (Is there a utility theory resolution to the Allais-paradox?). Journal of Economic Theory, 46, 145-153.

Simon, H. A. (1956). Rational choice and the structure of the environment. Psychological review, 63(2), 129.

Svenson, O. (1979). Process descriptions of decision making. Organizational behavior and human performance, 23(1), 86-112.

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

Tukey, J. (1980). We need both exploratory and confirmatory. The American Statistician, 34, 23-25.

Tversky, A. (1967). Additivity, utility, and subjective probability. Journal of Mathematical psychology, 4(2), 175-201.

Tversky, A. (1969). Intransitivity of preferences. Psychological Review, 76(1), 31-48.

von Neumann, J & Morgenstern, O. (1944). The theory of games and economic behavior. Princeton University Press.

Appendix 1: Six replications of Tversky's (1969) lottery experiment

Earlier replications

Three early replications were reviewed by Regenwetter et al. (2011) and by Cavagnaro and Davis-Stober (2014): those by Montgomery (1977) and Ranyard (1977); and the partial replication by Tsai and Böckenholt (2006).

In Montgomery's (1977) replication, 21 Swedish students undertook the same pre-test as that used by Tversky, choosing from a variation of the original lottery set in which the payoffs were from 40 to 50 SEK in steps of 2.50 SEK. Seven participants met Tversky's selection criterion, but because of resource limitations only five participated in Montgomery's main study. This was resource intensive since think aloud protocols as well as choice data were elicited and analyzed. The five selected participants undertook a single session in which each pair from the lottery set was presented ten times, interspersed with filler trials. Montgomery found that the choice data from all of them violated WST, although no significance test was applied. Regenwetter et al. (2011) reanalyzed the data and reported (in their online supplement) that choices from four participants significantly violated the TI condition at the 5% level or lower. However, Cavagnaro and Davis-Stober (2014) found strong evidence against transitivity for all five participants.

Ranyard (1977) carried out two relevant experiments with British students (N = 8; 20) in which the lottery payoffs were also changed to local currency, from 80 to 100 pence in steps of 5 (100 pence = 1 GBP)[5]. In the first, in individual sessions for each participant, all pairs from the

[5] Since the second experiment did not replicate Tversky's (1969) experiment it will not be reviewed here. Regenwetter et al. (2011) and Cavagnaro & Davis-Stober (214) reanalyzed it erroneously as if it were a replication of Tversky's.

lottery set were presented with eight replications, interspersed with filler trials. The lottery display format was changed, with both S and P information presented numerically. Tversky (1969) had chosen to present the latter in analogical form as sectors of a circle (representing a wheel of fortune with a spinner that could fall in either win or not win sectors) with their sizes determining probabilities of outcome. This was to induce an LS-type strategy in which small differences in S, presented numerically, were clearer than small differences in P. However, Ranyard argued that small sector differences may have been difficult to discriminate and consequently, intransitive cycles of choice may have to some extent been perceptual rather than preferential phenomena. He therefore changed the lottery format, displaying values on both dimensions clearly as numbers, in order to test whether Tversky's findings could be replicated when perceptual factors were controlled. After the main session, participants were asked how they had chosen from the target lottery pairs. They were classified as LS if they gave a "spontaneous and unambiguous statement that the basis for choice switched from probability to payoff as probability differences changed" (Ranyard, 1977, p. 454). On this basis two of the eight participants were classified in the LS group. For each choice pattern a measure of degree of departure from a transitive order, known as Slater's i, was calculated (Ranyard, 1976). This is the minimum number of choices that must be changed to produce a transitive order. As predicted, the median value of this statistic was significantly higher in the LS compared to the non-LS group. It should be noted that Ranyard (1977) analyzed the data from all participants but presented the individual data of the LS group in order to show how their choices compared to Tversky's. Regenwetter et al.'s (2011) reanalysis of the two LS participants of Ranyard's experiment 1, reported in their online supplement, found that one of these two choice patterns falsified the linear order mixture model by significantly violating the TI condition at the $p < .05$ level. However, Cavagnaro and Davis-Stober (2014) found strong evidence against transitivity in both cases.

The partial replication by Tsai and Böckenholt (2006) differed from Tversky's (1969) study in several respects: (1) only four of the five lotteries were presented (a,b,c and d in Table 1), with payoffs multiplied by four to take account of inflation; (2) each of the six pair comparisons of these lotteries were presented 120 times across several sessions per day for five days; (3) the five student participants were not screened using Tversky's pre-test procedure; and (4) unlike the other replications, they were motivated with a payment not dependent on their lottery choices[6]. One aim of the study was to compare an independent response (IR) with dependent response (DR) models. DR models assume that the utility of a lottery is dependent of the item with which it is compared, whereas IR models assume that they are independent. The authors did not analyze the overall binary choice proportions, focusing instead on the correlations between choices across pairs presented. They concluded that the pattern of correlations that they observed were consistent with a change in choice strategy for pairs of adjacent lotteries (i.e. (a,b), (b,c) etc.) compared to non-adjacent pairs. Despite these findings supporting Tversky's general position, Regenwetter et al.'s (2011) reanalysis of the five participants' overall binary choice proportions found that the data were consistent with their transitive preference model, since no significant violations of the TI condition were found. Cavagnaro and Davis-Stober's (2014) analysis confirmed this.

[6] We are grateful to Ulf Böckenholt for clarifying these methodological details in a personal communication.

Regenwetter et al.'s (2011) replication

Regenwetter et al. (2011) replicated and extended Tversky's (1969) lottery experiment with 18 US student participants. In contrast to Tversky' experiment, no pre-test was given and all carried out the main experiment. Three lottery sets were presented: (1) Cash 1, current price equivalents of Tversky's original lotteries, from \$22.40 to \$28.00 in steps of \$2.60, combined with the original P values; (2) Cash 2, similar P values (.28 to .44 in steps of .04) with S values varied so that the lotteries had equal expected value (\$20.00, \$22.00, \$24.44, \$27.50 and \$31.43); (3) a set (not discussed further here[7]) with non-monetary prizes of equal cash value, such as about 15 sandwiches. In a single session for each participant, the ten lottery pairs from all three sets were presented 20 times, with filler trials interspersed (818 choices in total). Rather than weak stochastic transitivity, the triangle inequalities condition was tested. Regenwetter et al. reported that in the first lottery set the linear order mixture model could be rejected for only one choice pattern because triangle inequalities were significantly violated ($p < .01$). The same was true for the second lottery set, and interestingly, it was the same participant. In their 2010 paper, the authors presented a comparison of violations of TI and WST on their data. This found that WST was violated significantly by one participant in the Cash 1 set and by two different

[7] This data was excluded because the lottery set departs in a non-trivial way from Tversky's original paradigm. On the positive side, the noncash payoffs were real lottery outcomes (unlike Tversky's (1969) experiment 2 choice alternatives which were hypothetical). However, they do not have a clear dimensional structure, unlike money which is a single quantitative dimension. Consequently, the NLAD strategy is not readily applicable in this context. Furthermore, the observed choice behaviour in this set was quite different to that in the cash sets: for many participants the choice seems to have been very strongly determined by the noncash prizes, and there is some indication that preferences in the binary choice task were reversed for some prizes compared to preferences in the prior ranking task. If explored further, this may turn out to be a preference reversal phenomenon, or due to the ranking procedure being unreliable.

participants in the Cash 2 set. Cavagnaro and Davis-Stober's (2014) similarly found strong evidence against transitivity for only two participants in each of sets Cash 1 and 2.

Cavagnaro & Davis-Stober (2014): a review, a replication and an extension

This paper follows up Regenwetter et al. (2011) by carrying out a more extensive analysis of previous replications of Tversky's (1969) experiment 1, and by reporting a new replication and extension with 29 student participants. This experiment had a 2 x 2 factorial, repeated-measures design with one factor being lottery set (set 1 versus set 2), and the other time pressure (with versus without). Thus, each participant carried out the task twice with each lottery set, once under time pressure and once with no time pressure. Lottery set 1 was that used by Tversky, with win amounts adjusted to take account of inflation. Set 2 had payoffs varied so that expected win increased with win amount, in contrast to set 1 in which it increased with probability of winning. The experiment was carried out in a single session, with all critical pairs presented 12 times to give 480 choices, interspersed with filler trials. The participants' standard payment was increased by the outcome of a randomly selected trial.

Whereas Regenwetter et al. (2011) focused on testing the goodness of fit of the mixture model of transitive preferences (MMPT), this paper carries out a comparative analysis of four models of transitive preference: MMPT, weak stochastic transitivity (WPT), moderate stochastic transitivity (MST) and strong stochastic transitivity (SST). These were compared with a baseline, encompassing model which permits intransitive preference, denoted M1. The authors adopt the Bayes factor, defined as 'the ratio of the marginal likelihoods of two models, derived from Bayesian updating', as the criterion for model evaluation. They argue that 'strong evidence in favor of M1 over any other [transitive] model means that the restrictions imposed by that model are not supported' (p.108). When comparing a transitive model against the baseline, a value greater than 3.16 defined strong evidence in favor of the transitive model, whereas a value less than .316 is taken as strong evidence against, and in favor of the baseline. When more than one transitive model has strong evidence in favor, the one with the highest value is considered to have the best fit. When all Bayes factors are between the above values the evidence is inconclusive. For set 1 without time pressure, which directly replicated Tversky's (1969) experiment, the authors found that although there was strong evidence in support of one or other of the transitive preference models for most participants (19/29), there was strong evidence in favor of M1, which allows intransitivity, in six cases (21 percent), while the evidence from four was inconclusive. The findings for the other three conditions were similar.

Kalenscher et al.'s (2010) replication

Kalenscher et al. (2010) extended previous replications of Tversky's (1969) lottery task, notably by measuring neural correlates of choice behavior. To this end, the 31 university-recruited Dutch participants carried out the task in an fRMI scanner. They were right-handed, had normal or corrected normal vision, and were screened to ensure they were fit to participate under these conditions. In addition to a 25 euro participation fee, they were paid the outcome of the lottery they chose on a randomly selected trial. Tversky's lottery set was modified such that probability of win was decimalized, varying from .29 in steps of .04 to .41, and presented graphically as a proportion of a vertical bar. In addition, the win amounts were presented in USD, from 400 in steps of 25 to 500, although the actual win was paid in euro converted at a rate 100 USD to one euro. Each pair from the lottery set was presented 20 times. These critical trials were interspersed with fillers in which one lottery dominated the other, since either the winning amounts or their probabilities were the same. This gave a total of 440 trails. There was a small number of missing choices because participants occasionally failed to respond within the 4s allowed.

The researchers measured degree of intransitivity with an index based on the number of preferences that had to be changed to eliminate violations of WST, which was specified a little differently to earlier, with $.4 \leq p(a,b) \leq .6$ defining $a \sim b$. The index was normalized by dividing the number of changed preferences by the maximum possible, so that it varied from 0 to 1. Participants were classified as intransitive if the index was $\geq .3$, benchmarked from a simulation of the choice behavior resulting from a stochastic version of the transitive expected utility model. Four participants were discarded from full analysis because of excessive head movements in the scanner, although choice data was accepted for three of them. For this sample of 30, 18 participants were classified as intransitive and 12 as transitive. Subsequently, however, Brown, Davis-Stober & Regenwetter (2015) applied Bayes factor analysis to this data and concluded that there was strong evidence of transitive preferences for a larger proportion, 14/30, and strong evidence of intransitivity for a smaller proportion, 7/30, with evidence from the remaining nine participants being inconclusive. Nevertheless, there was significant overlap in the two classifications, since all of Brown et al.'s intransitive participants' choices were classified the same by Kalenscher et al., and of Brown et al.'s 14 transitive participants, 9 were so classified by Kalenscher et al. Another interesting behavioral finding was that on average for intransitive participants the proportion of choice for the higher win probability was strongly related to dimension difference, whereas for transitive participants it was not.

Tables

Table 1. *Tversky's (1969) lottery set and proportions of row lottery choices over the column ones exhibiting violations of WST predicted by: (1) Lexicographic Semi-order heuristics (LS1, LS2); and (2) Non-linear additive difference strategies (SAD1, SAD2)*

Lotteries			Dimensional processing model									
			LS1					LS2				
P	S	Label	a	b	c	d	e	a	b	c	d	e
7/24	5.00	a	-	.8	.8	.2	.2	-	.2	.8	.8	.8
8/24	4.75	b		-	.8	.8	.2		-	.2	.8	.8
9/24	4.50	c			-	.8	.8			-	.2	.8
10/24	4.25	d				-	.8				-	.2
11/24	4.00	e					-					-
			SAD1					SAD2				
P	S	Label	a	b	c	d	e	a	b	c	d	e
7/24	5.00	a	-	.8	.5	.3	.1	-	.3	.5	.7	.9
8/24	4.75	b		-	.8	.5	.3		-	.3	.5	.7
9/24	4.50	c			-	.8	.5			-	.3	.5
10/24	4.25	d				-	.8				-	.3
11/24	4.00	e					-					-

Note.

(1) the five lotteries of the set have the same form, win S dollars with probability P, otherwise win nothing, e.g. in lottery a, $5.00 can be won with probability 7/24.

Table 2. *Tversky's (1969) lottery set and* observed proportions of row lottery choices (over the column ones) for two of his participants

Lotteries			Participant									
			4[1]					8				
P	S	Label	a	b	c	d	e	a	b	c	d	e
7/24	5.00	a	-	.50	.45	.20	.05	-	.60	.70	.75	.85
8/24	4.75	b		-	.65	.35	.10		-	.65	.75	.85
9/24	4.50	c				.70	.40			-	.60	.80
10/24	4.25	d				-	.85				-	.40
11/24	4.00	e										-

Note: [1]This participant's choice pattern significantly violated weak stochastic transitivity on Tversky's (1969) test, although Regenwetter et al. (2011) found that it did not significantly violate the triangle inequalities condition.

Table 3. Reanalysis of Tversky (1969): descriptive statistics, ML parameter estimation and goodness of fit (-2lnLL) of the models M_0, SAD, CCP, and RC (number of parameters in parentheses)

Part.	mean cp	CI	N out.	B	N WST	a_0	a_1	M_0(10)	SAD (2)	CCP (1)	RC (0)	Category
1	.62	.40-.80	3	-0.21	0,2,1	2.52	-0.98	31.75	**33.73**	72.44	84.07	SAD INT
2	.53	.35-.75	5	-0.12	1,1,0	1.16	-0.52	33.46	**44.42**	57.19	57.91	SAD INT
3	.69	.50-.85	4	-0.20	0,1,1	2.94	-1.00	28.38	**40.65**	78.49	108.12	SAD INT
4	.43	.25-.65	5	-0.23	3,2,1	1.97	-1.22	30.77	**37.98**	87.24	91.76	SAD INT
5	.60	.40-.80	2	-0.10	0,2,0	1.30	-0.44	33.27	**43.87***	53.06	61.11	NM
6	.72	.55-.90	2	-0.18	0,0,1	2.97	-0.95	28.13	**42.84**	76.42	114.64	SAD INT
7	.59	.40-.80	0	0.02	1,0,0	0.20	0.08	33.93	42.68	**43.02**	49.53	CCP(O)
8	.70	.50-.85	5	0.10	0,0,0	-0.22	0.55	32.08	**37.07**	48.19	79.43	SAD WST

Notes:
(1) Part., participant number

(2) Mean cp, mean choice proportion of the row lottery chosen over the column lottery.
(3) CI, confidence interval, the lower and upper bound of choice proportion for 95% CI.
(4) N outliers, number of choice frequencies outside the CI for the mean cp.
(5) B is the slope of the regression equation (B unit change in cp for 1 unit change in level of difference) nb: if B = .08, for every change difference of 1, choice proportion changes by 7.5%.
(6) N WST, number of triads out of three (a,b,c), (b,c,d), (c,d,e); number of 4-tples out of two (a,b,c,d), (b,c,d,e), and number of 5-tples out of one (a,b,c,d,e) violating WST in a manner predicted by Tversky (1969).
(7) -2LL values in bold-italics indicate the best fitting model which is: (1) a good fit, i.e. the fit of the unrestricted model is not significantly better; and in relation to the other variants (2) is significantly better than models with fewer parameters; and (3) models with more parameters are not significantly better. (All based on likelihood ratio tests with degrees of freedom the difference in number of parameters of the two models tested, test statistic difference in -2LL of the two models, assumed to be distributed as chi-square).
(8) Where none of the three models is a good fit on the above criteria, the -2LL value underlined indicates the best fitting model of these (code in parentheses). In these cases the unconstrained model is considered the overall best fitting model (indicated by bold/italic font).

* For participant 5, while the SAD model is a good fit, the cubic model is significantly better, in this case indicating a nonmonotonic pattern of choice proportions.

Table 4. Reanalysis of earlier replications: Tsai and Böckenholt (2006) r = 120; Montgomery (1977) r = 10; Ranyard (1977) r = 8. Descriptive statistics, ML parameter estimation and goodness of fit (-2lnLL) of the models M_0, SAD, CCP and RC (number of parameters in parentheses)

Part.	mean cp	CI	N out.	B	N WST	a_0	a_1	M_0(10)	SAD (2)	CCP (1)	RC (0)	Category
Ts1	.50	.40 - .60	2	-.09	2,1	0.55	-0.34	31.33	***38.56***	49.73	49.73	SAD INT
Ts2	.49	.40 - .60	0	-.13	2,1	0.89	-0.55	31.22	***33.30***	62.17	62.26	SAD INT
Ts3	.53	.45 - .62	0	-.04	0,1	0.37	-0.16	31.41	33.66	36.23	***38.03***	RC
Ts4	.51	.40 - .60	1	-.12	1,1	0.85	-0.48	31.29	***33.32***	55.30	55.86	SAD INT
Ts5	.50	.40 - .60	0	-.04	1,1	0.27	-0.15	31.43	34.59	36.94	***36.97***	RC
M1	.68	.40 - .90	3	-.22	0,2,0	3.16	-1.13	21.25	***25.08***	44.09	54.10	SAD INT
M2	.20	.30 - .80	5	-.30	1,2,1	3.40	-1.69	18.55	***28.08***	61.69	61.89	SAD INT
M3	.52	.30 - .80	6	-.36	2,2,1	4.83	-2.46	14.50	***23.12***	72.96	73.16	SAD INT
M4	.45	.20 - .70	7	-.37	3,2,1	5.00	-2.96	13.21	***19.75***	80.14	80.94	SAD INT
M5	.48	.20 - .70	3	-.16	0,1,0	1.27	-0.64	***21.62***	<u>42.04</u>	49.50	49.50	(SAD INT)
R1	.59	.25 - .88	3	-.25	0,1,1	2.86	-1.23	20.29	***29.45***	51.60	54.06	SAD INT
R2	.36	.13 - .63	6	-.31	3,2,1	4.30	-3.00	13.26	***23.11***	72.31	78.43	SAD INT

Note: See Table 3 notes for explanations of column contents

Table 5. Reanalysis of Regenwetter et al. (2011) Cash 1, r = 20. Descriptive statistics, ML parameter estimation and goodness of fit (-2lnLL) of models M_0 (10), SAD (2), CCP (1), and RC (0) (number of parameters in parentheses)

Part.	mean cp	CI	N out.	B	N WST	a_0	a_1	M_0 (10)	SAD (2)	CCP (1)	RC (0)	Category
1	.32	.15-.50	5	-.17	2,1,0	1.05	-0.98	30.82	**39.11**	69.03	94.06	SAD INT
2	.82	.65-.95	2	.01	0,0,0	1.10	0.21	**26.76**	43.54	<u>44.80</u>	133.50	(CCP(S))
3	.01	.00-.05	0	-.01	0,0,0							CCP(P)
4	.42	.20-.60	8*	-.24	1,1,0	2.12	-1.30	**26.85**	<u>74.7</u>	129.00	133.52	(SAD INT)
5	.04	.00-.10	0	-.02	0,0,0							CCP(P)
6	.26	.10-.40	5	-.11	2,1,0	1.90	-1.84	24.51	**34.27**	90.25	140.40	SAD INT
7	.10	.00-.20	0	-.02	0,0,0	-1.90	-0.18	24.06	26.44	**26.98**	178.66	CCP(P)
8	.02	.00-.10	1	-.02	0,0,0							CCP(P)
9	.37	.20-.60	2	-.01	0,0,0	-0.51	-0.02	33.18	46.87	**46.89**	61.65	CCP(O)
10	.07	.00-.15	0	-.04	0,0,0							CCP(P)
11	.03	.00-.10	1	-.03	0,0,0							CCP(P)
12	.40	.20-.60	4	-.19	3,2,1	1.40	-0.96	31.28	**44.23**	77.51	85.57	SAD INT
13	.36	.15-.55	0	-.10	0,0,0	0.32	-0.47	32.54	**46.32**	55.39	71.28	SAD WST
14	.99	.95-1.00	0	.01	0,0,0							CCP(S)

15	.25	.10-.40	2	-.10	0,0,0	0.13	-0.65	30.58	**35.57**	48.28	98.44	SAD WST
16	.09	.00-.20	2	.07	0,0,0	-4.38	0.85	**15.35**	<u>36.23</u>	48.01	208.94	(SAD WST)
17	.59	.40-.80	1	-.11	1,1,1	1.35	-0.48	33.01	**45.03**	56.01	62.52	SAD INT
18	.34	.15-.55	1	-.10	0,0,0	0.29	-0.52	32.42	**37.90**	48.45	72.09	SAD WST

Note: See Table 3 notes for explanations of column contents

Table 6. Lottery set and proportions of row lottery choices over the column ones for some participants classified as SAD INT choice patterns: top panel, Tsai and Böckenholt (2006); bottom panel, Regenwetter et al. (2011)

Lotteries			Participant 1				Participant 2			
P	S	Label	a	b	c	d	a	b	c	d
7/24	20	a	-	.61	.47	.43	-	.53	43	.33
8/24	19	b		-	.52	.38		-	.63	.43
9/24	18	c			-	.58			-	.58
10/24	17	d				-				-

			12					17				
P	S	Label	a	b	c	d	e	a	b	c	d	e
7/24	28.00	a	-	.50	.40	.25	.15	-	.65	.50	.40	.40
8/24	26.60	b		-	.80	.45	.05		-	.90	.40	.55
9/24	25.20	c			-	.65	.20			-	.70	.75
10/24	23.80	d				-	.55				-	.65
11/24	22.40	e					-					-

Note.

(1) The four or five lotteries of the set have the same form, win S with probability P, otherwise win nothing, e.g. in lottery a (bottom panel), $28.00 can be won with probability 7/24.

Table 7. Reanalysis of Cavagnaro & Davis-Stober (2014) Set 1, no time pressure (r = 12): Descriptive statistics, ML parameter estimation and goodness of fit (-2lnLL) of the models M_0 (10), SAD (2), CCP (1) and RC (0) (number of parameters in parentheses)

Part.	mean cp	CI	N out.	B	N WST	a_0	a_1	M_0 (10)	SAD (2)	CCP (1)	RC (0)	Category
1	.26	1-6	1	-.11	0,0,0	0.17	-0.67	24.38	***32.61***	40.68	69.92	SAD WST
2	.55	4-10	2	-.33	2,2,1	2.42	-1.1	26.72	***30.46***	58.77	59.97	SAD INT
3	.38	2-8	8	-.30	1,1,0	3.54	-2.36	16.73	***43.42***	104.81	111.40	SAD INT
4	.20	0-5	3	-.20	1,0,0	4.99	-5.00	13.308	***13.01***	66.08	112.34	SAD WST
5	.07	0-2	1	-.07	0,0,0							CCP(P)
6	.08	0-2	0	-.01	0,0,0							CCP(P)
7	.14	0-4	2	-.12	0,0,0	0.93	-1.80	***13.84****	30.03	48.56	117.00	NM
8	.31	1-6	8	-.25	3,2,1	3.93	-3.41	14.85	***24.19***	84.32	111.49	SAD INT
9	.58	4-10	3	-.19	1,1,1	2.08	-0.86	***25.88***	<u>45.24</u>	63.99	67.34	(SAD INT)
10	.27	1-6	0	.03	0,0,0	-1.26	0.16	25.57	38.92	***39.30***	66.47	CCP(O)
11	.03	0-1	1	-.00	0,0,0							CCP(P)
12	.52	3-9	3	-.23	1,2,3	2.19	-1.06	26.65	***33.24***	59.46	59.76	SAD INT
13	.12	0-3	0	-.09	0,0,0	0.38	-1.55	15.88	***19.25***	32.37	112.27	SAD WST
14	.39	2-8	1	-.18	1,2,1	1.30	-0.93	25.59	***33.73***	52.42	58.10	SAD INT
15	.30	1-6	5	-.25	1,0,0	2.89	-2.34	***16.74***	<u>34.32</u>	84.00	103.74	(SAD INT)
16	.63	5-10	0	-.10	1,0,1	1.61	-0.57	27.97	***34.04***	43.04	48.71	SAD INT
17	.05	0-2	1	-.05	0,0,0							CCP(P)
18	.37	2-7	9	-.34	3,2,1	5.00	-3.46	13.05	***24.16***	117.26	124.84	SAD INT
19	.28	1-6	5	-.23	3,2,1	2.57	-2.19	18.86	***26.91***	71.19	94.49	SAD INT
20	.10	0-3	1	-.06	0,0,0	-0.71	-0.88	***13.13***	<u>30.17</u>	35.57	123.90	(SAD WST)

21	.02	0-1	0	-.02	0,0,0							CCP(P)
22	.05	0-2	1	-.05	0,0,0							CCP(P)
23	.01	0-1	0	-.01	0,0,0							CCP(P)
24	1.0	0-0	0	.00	0,0,0							CCP(S)
25	.44	2-8	3	-.18	2,2,3	1.42	-0.86	27.00	***37.20***	54.82	56.46	SAD INT
26	.54	3-9	6	-.28	0,2,1	2.99	-1.43	25.22	***29.39***	70.31	71.14	SAD INT
27	.17	0-4	2	-.14	0,0,0	1.96	-2.46	16.44	***17.75***	46.68	104.90	SAD WST
28	.36	2-7	0	-.04	0,0,0	-0.22	-0.19	28.40	33.54	***34.46***	44.22	CCP(O)
29	.31	1-6	6	-.28	2,1,0	4.36	-3.25	***13.16***	36.88	110.31	123.90	(SAD INT)

Note: See Table 3 notes for explanations of column contents

Table 8. Reanalysis of Kalenscher et al. (2010), r = 20: Descriptive statistics, ML parameter estimation and goodness of fit (-2lnLL) of the models M_0 (10), SAD (2), CCP (1), and RC (0) (number of parameters in parentheses)

Part.	mean cp	CI	N out.	B	N WST	a_0	a_1	M_0 (10)	SAD (2)	CCP (1)	RC (0)	Category
1	.46	6-14	6	-.38	3,2,1	2.93	-1.66	***27.22***	44.43	122.39	123.67	(SAD INT)
2	.68	10-17	6	-.41	0,1,1	4.87	-1.89	***20.84***	37.63	132.12	157.15	(SAD INT)
3	.94	17-20	0	-.01	0,0,0							CCP(S)
4	.40	4-12	9	-.45	3,2,1	5.00	-3.28	19.19	***32.17***	185.89	193.94	SAD INT
5	.66	9-17	3	-.16	0,1,1	1.77	-0.54	31.98	***45.48***	58.21	79.06	SAD INT
6	.31	3-10	6	-.33	3,2,1	3.26	-2.56	***21.27***	40.86	134.35	163.97	(SAD INT)
7	.69	10-17	6	-.37	0,1,1	4.42	-1.64	25.41	***31.98***	111.03	140.65	SAD INT
8	.44	5-13	2	.07	0,0,0	-0.67	0.20	***33.18****	57.17	59.21	62.59	NM
9	.78	12-19	1	-.21	0,0,1	3.30	-0.97	***27.25****	51.57	80.32	144.31	LS INT
10	.55	7-15	7	-.39	1,2,1	3.43	-1.61	26.773	***47.27***	127.48	129.90	SAD INT
11	.06	0-3	0	-.01	0,0,0							CCP(P)
12	.74	12-19	2	-.22	0,0,1	3.87	-1.26	***25.94****	43.00	94.14	142.18	LS INT
13	.77	12-19	3	-.24	0,0,1	4.80	-1.54	***22.49****	39.82	104.14	165.69	LS INT
14	.54	7-15	3	-.20	1,0,0	1.46	-0.64	31.70	***57.79***	76.37	77.99	SAD INT
15	,950	17-20	1	.05	0,0,0							CCP(S)
16	.50	6-14	1	-.07	0,0,0	0.42	-0.22	***33.81****	49.04	51.47	51.49	NM
17	.51	7-15	4	-.18	2,2,0	1.20	-0.57	***32.45****	57.58	72.54	72.72	NM
18	.16	0-6	1	-.09	0,0,0	-0.54	-0.62	27.25	***30.86***	38.99	140.38	SAD WST
19	.63	8-16	3	-.19	0,1,1	1.80	-0.62	31.72	***45.78***	62.63	76.31	SAD INT

20	.77	12-19	6	-.37	0,1,1	5.00	-1.65	18.84	***32.61***	119.64	181.18	SAD INT
21	.59	7-15	1	.03	0,0,0	0.18	0.08	***33.64****	48.57	48.90	54.71	NM
22	.39	4-12	5	-.26	2,1,0	1.58	-1.11	31.34	***38.80***	79.10	89.77	SAD INT
23	.29	3-10	3	-.14	0,0,0	0.20	-0.57	31.36	***41.57***	52.95	87.58	SAD WST
24	.57	7-15	3	-.15	1,0,0	1.21	-0.46	***32.28***	60.50	70.42	74.36	(SAD INT)
25	.47	6-14	3	-.32	3,2,1	2.19	-1.22	29.47	***41.54***	93.52	94.50	SAD INT
26	.75	12-19	3	-.05	0,0,0	1.53	-0.21	***28.00***	58.52	60.20	112.52	(CCP(O))
27	.07	0-3	0	-.05	0,0,0							CCP(P)
28	.26	3-10	6	-.21	1,0,0	2.14	-2.00	23.82	***36.53***	98.67	146.71	SAD INT
29	.22	2-9	3	-.14	0,0,0	0.63	-1.12	26.84	***32.78***	58.79	127.84	SAD WST
30	.78	12-19	2	.04	0,0,1	1.79	-0.26	***26.84***	60.09	62.40	128.89	(CCP(O))

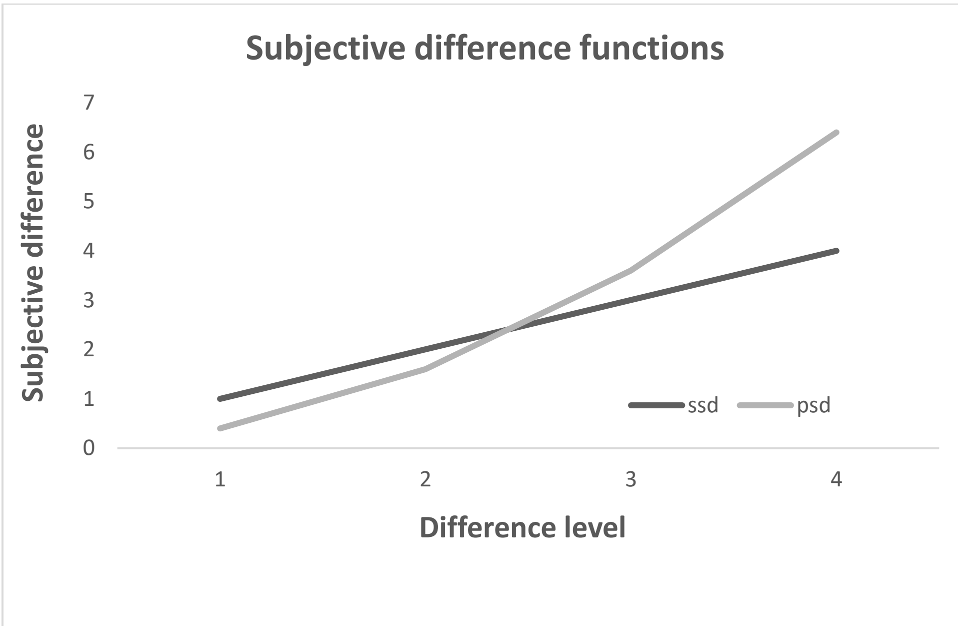
Note: See Table 3 notes for explanations of column contents

* For the seven participants asterisked in the M_0 column the cubic model was a good, and the best, fit. Inspection of the configuration of choice proportions revealed that intransitive preferences consistent with the LS heuristic were observed in three cases (classified as LS INT) and transitive preferences with a nonmonotonic relationship between choice proportion and dimension difference were observed in four cases (classified as NM).

Table 9. Percent of participants in each category (seven studies, total N = 129).

Study	**Category**						**N**
	SAD/LS INT	**SAD WST**	**CCP**	**RC**	**M_0/NM**	**UNKNOWN**	
Tversky (1969)	27.8	5.6	5.6		5.6	55.6	18
Montgomery (1977)	19.0				4.8	76.1	21
Ranyard (1977)	25.0					75.0	8
Tsai & Böckenholt (2006)	60.0			40.0			5
Regenwetter et al. (2011)	22.2	16.7	44.4		16.7		18
Cavagnaro & Davis-Stober (2014)	34.5	13.8	34.5		17.2		29
Kalenscher et al. (2010)	46.7	6.7	13.3		33.3		30
Frequency total	42	10	23	2	20	32	129
Percent	32.6	7.8	17.8	1.6	15.5	24.8	

Figures



.

Figure 1. Example dimension subjective differences as a function of difference level: linear for the payoff dimension, $\varphi_s(\delta) = \delta$, and nonlinear for the probability dimension, $\varphi_{p}(\delta) = .4\delta^2$.

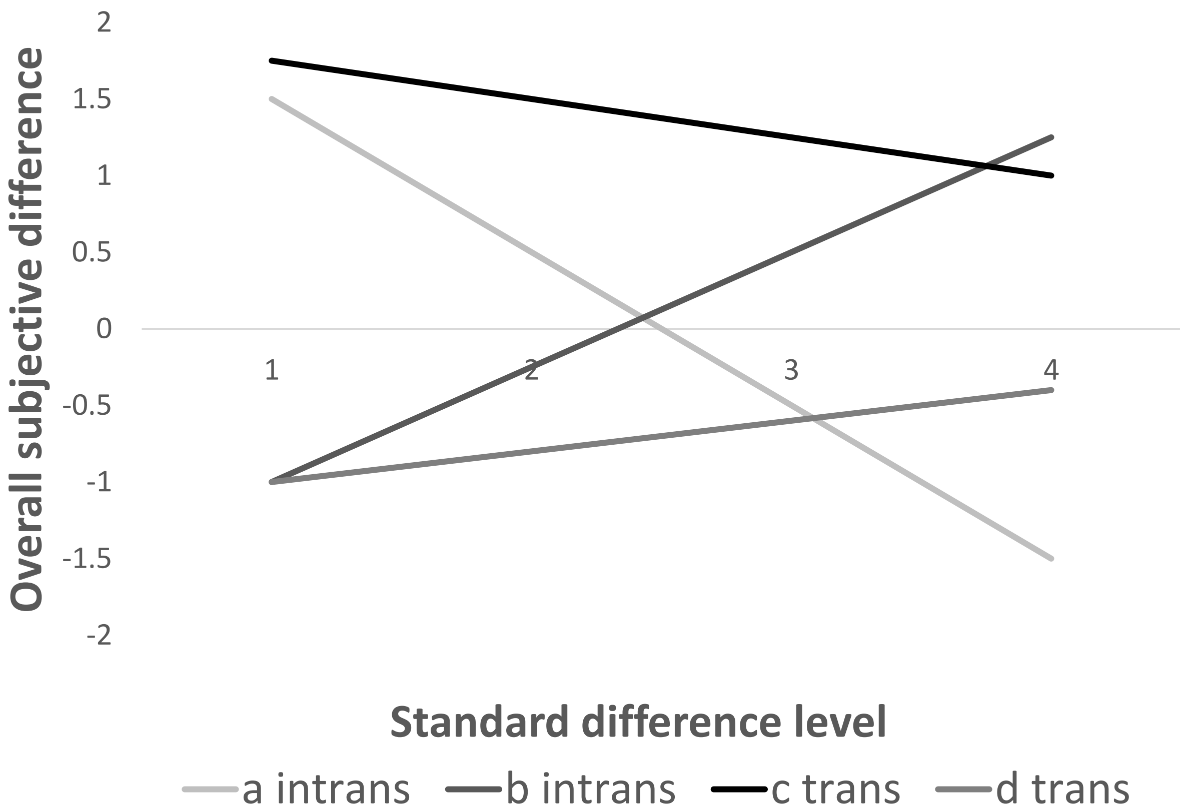


Figure 2. Overall subjective difference functions of the SAD model predicting intransitive preferences (a, b) or transitive preferences (c, d).

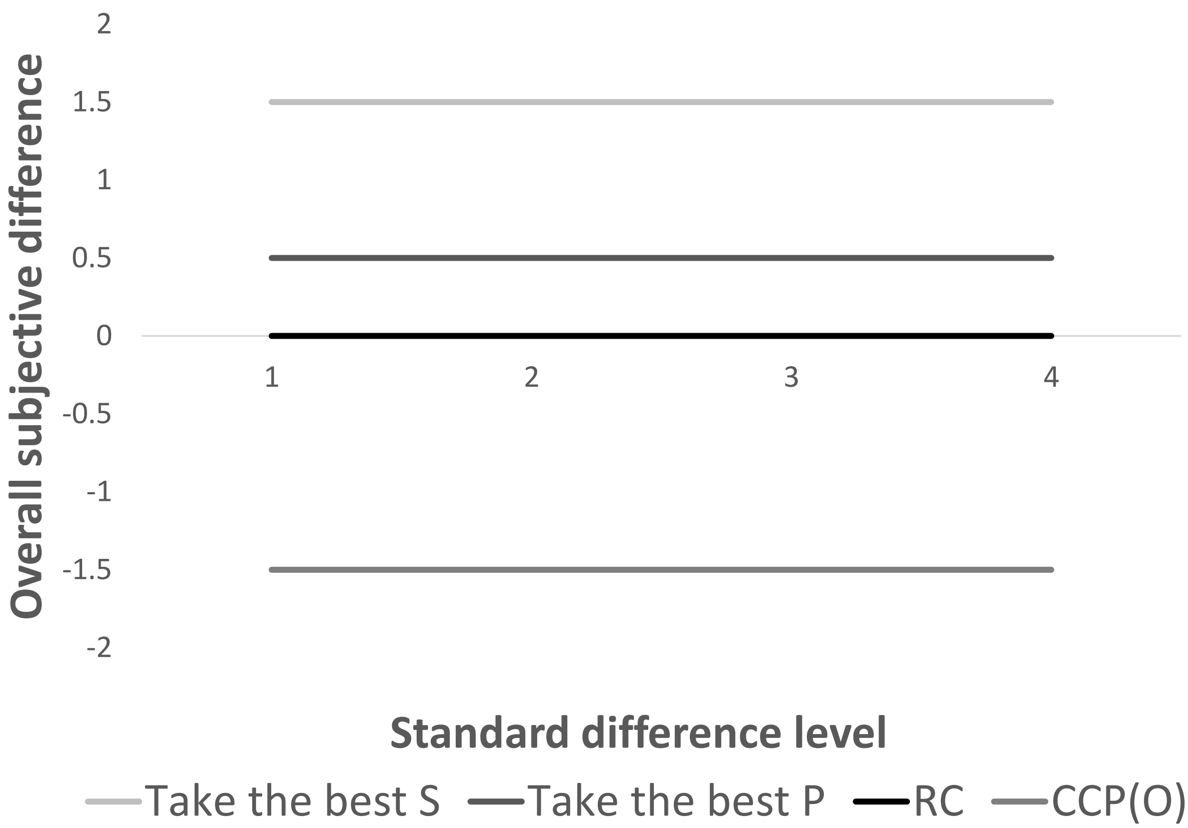


Figure 3. Overall subjective difference functions for three versions of the CCP model predicting transitive preferences, and the random choice model.