Title page with All Author Information

Testing a Class of Models that Includes Majority Rule and Regret Theories: Transitivity, Recycling, and Restricted Branch Independence

Michael H. Birnbaum

Department of Psychology, California State University, Fullerton, CA 92834, USA

Enrico Diecidue

INSEAD, Fontainebleau, France

Short Title: Testing Majority Rule and Regret Theory

Filename: Majority_Rule_Regret_49

Supplemental materials: http://????

Prod. Ed: Our supplementary materials are needed to understand the article and therefore, they should be copy-edited, according to APA Website. These tables are cited in the paper with terminology such as "Table SM.1", etc. They appear at the end of this manuscript, after the figures.

Contact Info: Mailing address:

Prof. Michael H. Birnbaum,
Department of Psychology, CSUF H-830M
P.O. Box 6846,
Fullerton, CA 92834-6846

Email address: mbirnbaum@fullerton.edu

Phone: 657-278-2102

Acknowledgments: Support was received from National Science Foundation Grants, BCS-0129453 and SES DRMS-0721126. Financial support was also provided by research grant CONCORT FP7-PEOPLE-2011-ITN. We thank Jeffrey P. Bahra, Roman Gutierrez, Ngo Huong, and Adam LaCroix for assistance with data collection. We thank Han Bleichrodt, Daniel Cavagnaro, Jonathan Leland, R. Duncan Luce, and anonymous reviewers for useful suggestions and discussions of these ideas.

Abstract

Six experiments compared four classes of decision-making models that make different predictions for tests of three diagnostic behavioral properties. These properties are transitivity of preference, recycling of intransitivity, and restricted branch independence. Two experiments tested decisions based on advice from friends, a situation in which majority rule was previously thought to apply. Both experiments rejected the hypothesis that more than a small percentage of participants might have used majority rule. Four experiments with choices between gambles tested a very general family of models that includes majority rule and regret theory as special cases. Such models not only allow but in certain cases require intransitive preferences and recycling, and they require that restricted branch independence must be satisfied. Two experiments used regret theory fitted to previous data to predict where to find violations of transitivity and recycling; no evidence was found to confirm these predictions; instead, a few participants showed the opposite pattern of intransitivity and recycling from that predicted. Two experiments with monetary incentives used a new experimental design in which a general model including regret theory, salience weighted contrasts, and perceived relative arguments (PRAM) must violate transitivity. Most participants did not show the predicted response patterns required by these models. Combining results of six studies, we conclude that the family of models that includes regret theory, PRAM, and majority rule is not an adequate descriptive model of how most people make decisions. Transitive, configural weight models best described most participants' data.

Keywords: branch independence, choice theory, decision-making, independence, recycling intransitivity, regret theory, transitivity of preference, utility theory

Introduction

How do people make choices when deciding between two alternatives? In this article, we compare theories of the following two tasks:

1. Suppose you are planning to take a vacation in one of two cities, and three equally credible friends, *A*, *B*, and *C*, give their ratings (opinions) as to how well they think you would enjoy each city, as shown in Figure 1. Based on their ratings, which city would you rather visit?

Insert Figure 1 about here.

2. Suppose there is an urn containing an equal number of Red, White, and Blue marbles. Suppose you will win a cash prize according to the matrix of payoffs shown in Figure 2, depending on your choice of which gamble to play and the color of marble that will be drawn randomly from the urn. Which gamble would you rather play?

Insert Figure 2 about here.

Many models have been proposed to describe how people make such choices. These models can be categorized into four classes based on two fundamental behavioral properties: transitivity and restricted branch independence.

If preferences are *transitive*, it means that if a person prefers option *F* to *G* and *G* to *H*, then that person should prefer *F* to *H*, apart from random error. *Branch independence* is the assumption that if two options share some attribute in common, then that common attribute should have no effect on the decision induced by other attributes, apart from error. In decision-making, this property is required by Savage's (1954) "sure thing" axiom. Restricted branch independence is a weaker (and "purer") form of this axiom, which is specified more precisely below.

A number of models of decision-making either satisfy or violate transitivity and either satisfy or violate restricted branch independence. By testing these two properties we can determine

which models are viable as empirical descriptions of behavior and which are not.

Models of Decision Making and their Properties

Figures 3 and 4 depict two broad classes of models that are transitive or intransitive, respectively. Let $F = (x_F, y_F, z_F)$ represent a city that has been evaluated by three friends, where x_F, y_F, and z_F are ratings of City F by friends A, B, and C, respectively; similarly, $G = (x_G, y_G, z_G)$ is another city where x_G, y_G, and z_G are ratings of City G by the same three friends, respectively. [Alternately, F and G might represent gambles in which (x_F, y_F, z_F) are the monetary prizes to be won if a marble drawn randomly from the urn were Red, White, or Blue, respectively when Gamble F was chosen, and x_G, y_G, and z_G are the respective prizes for Red, White, or Blue when Gamble G was chosen.]

Figure 3 depicts an outline of processes in a family of transitive models. Note that each rating (or monetary prize) has a subjective value. The mapping of objective into subjective values, $u(x)$, is called the *valuation* stage (denoted V in the figure). These are *integrated* (in stage I) within each city to form separate evaluations of the cities, $U(F)$ and $U(G)$, which are *compared* (in stage C) to form a subjective preference, which is mapped in the *judgment* or response production stage (J), to an overt choice response, R.

Insert Figures 3 and 4 about here.

Figure 4 shows the corresponding outline for a class of intransitive models. In this case, the person compares values of the attributes between cities (or compares the utilities of the monetary prizes between gambles) to form contrasts, $\psi(x_F, x_G)$; next these contrasts are integrated to form an overall preference. In the perceived relative arguments model (PRAM) of Loomes (2010), for

example, these contrasts are called *relative arguments* for F over G. As we see below, this family of models can (and in some cases must) violate transitivity with appropriately selected stimuli.

The small arrows pointing southwest in each figure depict possible loci for noise, variability, or random error in the processes of evaluation, integration, comparison, and judgment. More will be said about this topic in the section on true and error models.

Additive Contrast Models including Majority Rule Violate Transitivity

Let $F = (x_F, y_F, z_F)$ represent ratings of city F, where x_F, y_F, and z_F are evaluations by friends A, B, and C, respectively. Let $G = (x_G, y_G, z_G)$ represent ratings of City G by the same three sources. Define the *additive contrast model* as follows:

$$\delta(F, G) = \psi_1(x_F, x_G) + \psi_2(y_F, y_G) + \psi_3(z_F, z_G) \qquad (1)$$

where $\delta(F, G)$ is the subjective preference between F and G, and ψ_1, ψ_2, and ψ_3 are skew-symmetric contrast functions having the property, $\psi_\iota(x, y) = -\psi_\iota(y, x)$, hence $\psi_\iota(x, x) = 0$. The additive contrast model assumes $F \succ G \Leftrightarrow \delta(F, G) > 0$, where $F \succ G$ denotes that F is preferred to G.

The *majority rule* is a special case of Equation 1 in which the three functions, ψ_1, ψ_2, and ψ_3, are all identical, the same as ψ, where $\psi(x, x') = 1$, 0, or –1, when $u(x) > u(x')$, $u(x) = u(x')$, or $u(x) < u(x')$, respectively. If all three sources rate F better than G, $\delta(F, G) = 3$; if two of the three sources rate F better than G and one rates G better, $\delta(F, G) = 1$. Thus, as long as a majority of sources give higher ratings to F than G, people should prefer F to G, no matter what those ratings are.

The additive contrast model is not transitive, nor is majority rule. For example, $F = (2, 6, 7) \succ G = (4, 5, 5)$ and $G = (4, 5, 5) \succ H = (3, 7, 3)$, yet $H = (3, 7, 3) \succ F = (2, 6, 7)$, violating transitivity, because in each choice problem, the preferred alternative had higher ratings by two of the three sources. Majority rule must violate transitivity with these stimuli, apart from random error, because this model has no free parameters. The more general, additive contrast model (Equation 1), however, allows either transitive or intransitive preferences for cities *F*, *G*, and *H*, but as we see in Experiments 5 and 6, it is possible to design a study in which intransitivity must occur according to the broadest generalization of majority rule via Equation 1.

Recycling

Let $F = (4, 5, 6)$; then we say that $F' = (6, 4, 5)$ is a *permutation* of *F* because it has the same three ratings of the cities (or monetary prizes), but these components may be differently associated with the friends (or the events). *Recycling* is the property that there exists a set of alternatives, *F*, *G*, and *H,* which form an intransitive preference cycle that can be reversed (producing the opposite intransitive cycle) by a suitable permutation of the component values. That is, $G \succ F$, $H \succ G$, yet $F \succ H$; which can be reversed, $F' \succ G'$ and $G' \succ H'$, yet $H' \succ F'$, where G', F' and H' are permutations of *G*, *F*, and *H*, respectively.

With the assumption that the contrast functions (ψ_1, ψ_2, and ψ_3) are all the same, Expression 1, including majority rule, implies recycling. For example, consider $F = (4, 5, 6)$, $G = (5, 7, 3)$, and $H = (9, 1, 5)$. According to majority rule, $G \succ F$ and $H \succ G$, but $F \succ H$. Now consider $F' = (6, 4, 5)$, $G' = (5, 7, 3)$, and $H' = (1, 5, 9)$. In this case, $G' = G$, and F' is the same as *F*, except the friends' ratings of the cities have been permuted across friends; similarly, H' is a permutation of

H. In this case, majority rule implies the opposite intransitive cycle, $F' \succ G'$ and $G' \succ H'$, but $H' \succ F'$.

Recycling requires violations of *permutation independence* but it requires more. Permutation independence is the assumption that $A = (x, y, z) \succ B = (t, u, v)$ iff $A' \succ B'$ where A' and B' are any permutations of A and B. Violations of permutation independence are sometimes called "juxtaposition" effects (Starmer & Sugden, 1993). Violations of permutation independence might be compatible with transitive models (as noted below), but recycling can only occur in intransitive models, because recycling requires a reversal of an intransitive cycle (opposite intransitive cycles) via permutation of the consequences. We are not aware of any previous theoretical or empirical treatment of recycling, which is tested in Experiments 2-4 and 6.

Restricted Branch Independence

Equation 1 satisfies *restricted branch independence* (RBI), which requires

$$S = (x, y, z) \succ R = (x, y', z') \Leftrightarrow S' = (x', y, z) \succ R' = (x', y', z') \qquad (2)$$

The basic idea is that the preference between S and R does not depend on any component that is the same in both alternatives. In Expression 2, S and R have a common branch (the first source's rating with the same value of x), whereas S' and R' share a common branch (first source's rating is now x' in both alternatives)*;* that is, the first source gave the same rating in both the SR and $S'R'$ choice problems. Similarly, if either of the other sources were to give the same evaluation to both cities, that evaluation should not affect the choice response either, aside from error.

Proof: According to Equation 1, $\delta(S, R) = \psi_1(x, x) + \psi_2(y, y') + \psi_3(z, z')$; because $\psi_1(x, x) = \psi_1(x', x') = 0$, $\delta(S, R) = \delta(S', R')$; hence, $S \succ R \Leftrightarrow S' \succ R'$. The proofs for the cases of common

ratings by the second or third sources work exactly the same way, since $\psi_2(y, y) = \psi_2(y', y') = 0$ and $\psi_3(z, z) = \psi_3(z', z') = 0$. In some applications, as in this article, RBI can also be called *joint independence* (e.g., see Birnbaum & Zimmermann, 1998).

Integrative Models Satisfy Transitivity

Now consider the family of *transitive, integrative models*:

$$U(F) = I[u_1(x), u_2(y), u_3(z)] \qquad (3)$$

Where $U(F)$ is the overall evaluation of city F, u_1, u_2, and u_3 are functions of the sources' ratings of the cities (or of the monetary prizes of gambles), and I is an integration function that combines information about the city (or gamble), as in Figure 3.

According to this family of models (Figure 3), a person prefers the city (or gamble) with the higher overall evaluation; i.e., $F \succ G \Leftrightarrow U(F) > U(G)$. Because the evaluations, $U(F)$, are numerical (and the relation, >, on real numbers is transitive), this family of models implies transitivity of preference, aside from random errors.

One model for the integration function (I in Figure 3) is the constant-weight averaging model (Anderson, 1974):

$$U(F) = [au(x) + bu(y) + cu(z)]/(a + b + c) \qquad (4)$$

where $u(x)$ is the subjective value of a rating of x, and a, b, and c are weights for the sources. The additive model, $U(F) = au(x) + bu(y) + cu(z)$, is equivalent to this averaging model when the sum of the weights is constant. When the number or credibility of sources of information is varied, the sum of the weights is not constant; in this case, the model also uses an initial impression, $u(x_0)$, which represents the judge's impression of the city in the absence of information from any sources

(Anderson, 1974; Birnbaum & Stegner, 1979). In this study, however, we do not manipulate those factors, so additive and constant-weight averaging models are equivalent in these studies, and the initial impression is not needed. Anderson (1974) noted that the weights of sources often depend on serial position.

When $a = b = c$, and when $u(x)$ is the identity function, we have a simple average value or expected value model. According to this simple version of averaging, preferences among cities should show no "risk aversion"; for example, people should prefer $L = (1, 6, 9)$ to $M = (5, 5, 5)$, because L has the higher average rating. However, people might avoid a "risky" option such as L to which one friend gave such a low rating, which would be consistent with Equation 4 if u is not linear.

Similarly, the simple average or expected value model implies there should be no effect of permutation of ratings across the sources. However, Equation 4 allows that people might assign more weight to the first friend than to the second, so it is possible that $L = (9, 1, 4) \succ M = (1, 9, 5)$ and yet $L' = (1, 9, 4) \prec M' = (9, 1, 5)$, which could be consistent with Equation 4 if $a > b > c$. (For example, let $u(x) = x$, $a = 3$, $b = 2$, and $c = 1$). Therefore, when these weights are not equal, Equation 4 can violate permutation independence ("juxtaposition" effects). But because this model is transitive, the issue of recycling is moot.

It is important to realize that although evidence of risk aversion and of violations of permutation independence may rule out the simple averaging model, these phenomena remain compatible with the constant-weight averaging model in its more general form of Equation 4. As noted in our Discussion of Experiments 1 and 2, previous research on the majority rule (Zhang, et al., 2006) did not allow the averaging model these standard features, so data compatible with

averaging models were uncritically accepted as evidence for majority rule.

Like the additive contrast model (Equation 1), additive and constant weight averaging models (Equation 4) imply RBI (see proof in Birnbaum & Zimmermann, 1998), which is violated by configural weight averaging models, discussed next.

Configural Weight Averaging Models

The configural weight averaging model (Birnbaum, 1974; Birnbaum & Stegner, 1979; Birnbaum & Zimmermann, 1998) allows weights of the components to depend on the configuration of attribute values. In the following model, weights depend on the ranks of the stimuli:

$$U(F)= [w_L\min(X, Y, Z)+w_M\text{med}(X, Y, Z) + w_H\max(X, Y, Z)] /[w_L+w_M+w_H] \qquad (5)$$

where min[X, Y, Z], med[X, Y, Z], and max[X, Y, Z] are the minimum, median, and maximum of the subjective values of the weighted evaluations, $X = au(x)$, $Y = bu(y)$, and $Z = cu(z)$, respectively; $a + b + c = 1$; and w_L, w_M, and w_H are the corresponding weights. When $w_L = w_M = w_H$, this model reduces to a constant weight averaging model (Equation 4). Note that whereas a, b, and c represent weights that might depend on serial positions and credibility of the sources, w_L, w_M, and w_H represent the effects of the relative positions of the source's opinions relative to the other opinions expressed for that same city; these weights are thus called "configural" because they depend on the relations among the stimuli (Birnbaum, 1974).

The configural weight model of Equation 5 is a special case of Equation 3, so it satisfies transitivity; however, it violates RBI. For example, suppose people give greater weight to lower evaluations, such that w_L, = 0.46, w_M = 0.33, and w_H = 0.21, that $a = b = c = 1$, and suppose $u(x) = x$. Let S = (1, 4, 6), R = (1, 2, 9), S' = (10, 4, 6), and R' = (10, 2, 9). We have $U(S) = 3.04 > U(R) = 3.01$, but $U(S') = 5.92 < U(R') = 5.99$, so $S \succ R$ but $S' \prec R'$, violating RBI.

Critical Tests Distinguish Four Classes of Models

Table 1 shows how to distinguish four classes of descriptive decision models by testing both RBI and transitivity. A number of models fall in each category. Models that satisfy both RBI and transitivity (apart from error) include expected utility (EU), subjectively weighted utility (Edwards, 1954), the "stripped" version (Starmer & Sugden, 1993) of original prospect theory (Kahneman & Tversky, 1979), and the adding or constant weight averaging models (Anderson, 1974), including Equation 4. Furthermore, if people were to edit choice problems by cancelling any attribute that is the same in both alternatives, as was theorized in original prospect theory (Kahneman & Tversky, 1979), those people would satisfy RBI, apart from error.

Insert Table 1 about here.

The configural weight models, including Equation 5, satisfy transitivity and violate RBI (Birnbaum, 1974; 1982; Birnbaum & Beeghley, 1997; Birnbaum & Navarrete, 1998; Birnbaum & Stegner, 1979; Birnbaum & Sutton, 1992; Birnbaum & Zimmermann, 1998; Johnson & Busemeyer, 2005). Birnbaum's (1999a, 1999b, 2008a) rank affected multiplicative (RAM) and transfer of attention exchange (TAX) models fall in this class, as do other rank-affected configural models (Luce & Marley, 2005; Marley & Luce, 2005).

Cumulative prospect theory (CPT) (Tversky & Kahneman, 1992) and rank and sign dependent utility (RSDU) (Luce & Fishburn, 1991, 1995, Luce, 2000) are also configural weight models that satisfy transitivity and violate RBI. In these studies, they are also special cases of Equation 5. However, these models can be tested against models such as TAX, by means of other critical experiments where TAX and CPT, for example, make differential qualitative predictions. For a summary of such critical tests, including evidence that strongly refutes CPT and RSDU models, see Birnbaum (2008a).

Although CPT can violate RBI, it requires an inverse-S decumulative probability weighting function in order to describe the Allais paradoxes and other common findings. This weighting function implies the opposite pattern of violation from that predicted by Birnbaum & Stegner's (1979) model (see Birnbaum & McIntosh, 1996) or Birnbaum's special TAX model (Birnbaum, 2008a). This differential prediction will be tested in these studies.

The additive contrast models of Equation 1, the perceived relative arguments model (PRAM) by Loomes (2010), the similarity model (Leland, 1994; 1998), salience weighted utility (Leland & Schneider, 2014), stochastic difference model (Gonzalez-Vallejo, 2002), regret theory (Loomes & Sugden, 1982), majority rule (Russo & Dosher, 1983; Zhang, Hsee, & Xiao, 2006) and the most probable winner models (Blavatskyy, 2006) can (and in some cases, must) violate transitivity, but they satisfy RBI. Some of these models can (and in some cases must) show recycling in specially constructed experimental designs.

Models that can violate both transitivity and RBI include the lexicographic semiorder (Tversky, 1969) and priority heuristic (Brandstätter, Gigerenzer, & Hertwig, 2006). These models do not always violate these properties. The studies reported here were designed to test models in the upper right cell of Table 1, including majority rule and generalizations that include regret theory, but they were not designed to test the family of lexicographic semi-orders and the priority heuristic. For a summary of tests devised specifically to test those models (lower right cell of Table 1), which present strong empirical evidence against that class of models, see Birnbaum (2010) and Birnbaum and Bahra (2012b).

Recycling is the property that if intransitive cycles are found, then the cycles can be reversed by suitable permutations of component values. It is therefore a test that distinguishes among models that violate transitivity. Recycling is implied by additive contrast models (including

regret theory and majority rule) and by PRAM, but not by all intransitive models.

Response Errors in Critical Tests of Formal Properties

Because the same person may not always make the same response when the same choice problem is presented twice in the same block of trials, violations (or satisfactions) of a formal property like transitivity, RBI, or recycling might occur by random error. In order to distinguish true violations or satisfactions of such properties from those that might arise from error or noise we need a proper model to assess if any given rate of observed violations is real (Birnbaum, 2004, 2011, 2013; Birnbaum & Gutierrez, 2007; Lichtenstein & Slovic, 1971).

The family of true and error (TE) models, as developed recently by Birnbaum and colleagues has strong advantages over certain rival or older approaches. These older methods can be represented as unnecessarily restricted special cases of TE models, as cases where it is not possible to unambiguously distinguish error from true intention, or as cases in which unnecessarily restrictive and false assumptions could lead to erroneous conclusions.

An early form of the TE model was termed the "trembling hand" or "tremble" model because this form of the model assumed that the rate of error was the same for all choice problems, as if error arises only in the response production, or judgment stage, *J*, of Figures 3 and 4. Assuming that error rates are constant when they are not could lead to wrong conclusions concerning formal properties such as transitivity.

Sopher and Gigliotti (1993) allowed error rates to differ between choice problems, but their approach potentially used more parameters than degrees of freedom in the data (Birnbaum & Schmidt, 2008). This approach could therefore be criticized because it used the assumption of transitivity in order to construct the error rates that could be criticized as "fudge factors" devised to force transitivity onto otherwise intransitive data.

The current approach to fitting and testing TE models avoids these problems of the early work (Birnbaum & LaCroix, 2008). However, this approach requires the experimenter to present each choice problem to the same participant at least twice in each block of trials (session). Reversals of preference by the same person to the same choice problem presented twice in the same block of trials are used to estimate error rates. With this experimental feature, there are more degrees of freedom in the data than parameters to estimate, so the TE models are testable (Birnbaum & Schmidt, 2008).

Error rates might differ for different problems because some problems are more "difficult" and provide more opportunities to mis-read information, mis-remember it, or mis-evaluate it, which would produce errors in the valuation stage, *V*. Similarly, different choice problems may involve differential amounts of error at any of the other stages depicted in Figures 3 and 4, including the integration (*I*) and comparison (*C*) stages, as well as in the judgment stage. It is an empirical issue whether or not error rates are the same for all choice problems.

TE models have been applied both for the situation of data from a group of people (as in this study), called *group* True and Error Theory (*g*TET) as well as the situation where one person performs the same task repeatedly, called *individual* True and Error Theory (*i*TET). See Birnbaum & Bahra (2012a, 2012b). Both approaches use the same fundamental assumption: errors are revealed by reversals of preference to the same choice problem by the same person in the same block of trials.

TE models need not make any assumptions concerning transitivity. Thurstone's and Luce's choice models, in contrast, require transitivity in the absence of error (Luce, 1959; Birnbaum & Schmidt, 2008); consequently, these models are not appropriate when transitivity is to be tested rather than assumed. If we want to investigate transitivity objectively, we should not force our

analysis to use an error structure that assumes transitivity on an underlying metric continuum. These transitive metric models are special cases of a TE model, but these special cases may or may not be descriptive of actual data.

In early research on the issue of transitivity, researchers investigated marginal (binary) choice proportions and asked if these were consistent with properties defined on binary choice proportions such as weak stochastic transitivity (WST). According to WST, if $P(FG) > ½$ and $P(GH) > ½$ then $P(FH) > ½$. Unfortunately, if there is a mixture within a group of people, in which everyone is transitive but different people have different true preference orders, or if in an experiment with a single individual that person is always transitive but changes true preference orders during the study, then WST can be violated. Thus, violations of WST do not necessarily mean that anyone was truly intransitive.

Some have argued that one should test the triangle inequality instead, which requires $P(FG)+P(GH)+P(HF) \leq 2$. This property has the advantage over WST in that if everyone is always transitive, the property will be satisfied. However, the triangle inequality can be satisfied when there are mixtures including systematically intransitive patterns (Birnbaum, 2013). Therefore, satisfaction of the triangle inequality does not mean that people are using only transitive processes.

TE models have strong advantages over the random preference approach (aka random utility) as in Regenwetter, Dana, and Davis-Stober (2011) because they do not require the assumption of response independence. Response independence requires that the probability of a response pattern (a combination of choice responses) is the product of the binary choice probabilities composing the combination. Response independence has been empirically tested and rejected in several studies (Birnbaum, 2012; Birnbaum & Bahra, 2012a, 2012b), including in Birnbaum's (2012, 2013) reanalysis of the data of Regenwetter, et al. (2011). Although Cha, Choi,

Guo, Regenwetter, and Zwilling (2013) claimed that Birnbaum's (2012) tests of independence were dubious, that the Regenwetter, et al. (2011) data might satisfy independence, and that the TE models must also satisfy response independence or become meaningless, Birnbaum (2013) showed that their main arguments were not correct.

Response independence can occur as a special case in the TE models when there is only one true preference pattern, but in general independence will be violated in this model (Birnbaum, 2013). As shown in Birnbaum (2013), the false assumption of response independence in the approach of Regenwetter, et al. (2011) could easily lead to the wrong conclusion that transitivity is satisfied in cases where a TE model would correctly diagnose the data as containing systematic intransitivity.

Consequently, the methods of investigating transitivity and other formal properties advocated by Regenwetter, et al. (2011) as well as other methods that restrict the focus to binary choice proportions could lead to wrong conclusions when the observed data might represent a mixture of preference patterns. Although recent papers call such methods "state-of-the art", we disagree. If responses represent mixtures generated either by variability within or between persons, response independence can be violated, and methods that assume independence could lead to wrong conclusions. Therefore, those approaches to studying transitivity that restrict their focus to binary choice proportions (e.g., Regenwetter, et al., 2011; Regenwetter, et al., 2014) cannot be trusted to yield correct conclusions (Birnbaum, 2011, 2013).

TE models require more extensive experiments than would have sufficed in the past, but TE models capture more detail in the data, and they supply more information from the analysis (Birnbaum, 2011, 2013). In particular, they require one to present each choice problem at least twice in each block of trials to each participant, in order to properly estimate error rates. But in

return for this extra experimental work, these models allow one to estimate the distribution (relative frequencies) of different true response patterns in a mixture.

In the case of an investigation of transitivity, TE models allow one to draw inferences regarding the percentage of participants who might use different transitive or intransitive strategies (or in the case of an individual, to estimate the proportion of time that the individual used transitive or intransitive response patterns). Rather than providing merely a Yes/No statistical test of, "are these data transitive?" TE models provide estimates of the probabilities of different transitive and intransitive response patterns in the mixture. The basic TE model is presented in more detail in the Results of Experiments 1.

The rest of this paper is organized as follows: Experiments 1 and 2 present two studies testing majority rule as a model of how people decide what city to visit based on their friends' opinions of how much they would like to visit those cities. The results section of Experiment 1 introduces the TE model, shows how it is applied, how it can be tested, how it can be compared with models that predict response independence, and how it can be used to estimate the percentage of participants who had each preference pattern, including the percentage who satisfied or violated transitivity. Experiments 3 and 4 test predictions of regret theory for choices among gambles, as fit to previous data for intransitive cycles. This part of the paper shows how to analyze choices with respect to the parameters of a model and how to use parameters estimated in one study to design a test of a formal property in another study. Experiments 5 and 6 employ a highly restricted design in which a general class of models must violate intransitivity, apart from error.

Experiments 1 and 2: Decisions Based on Advice

According to majority rule, when people compare two alternatives, as in Figure 1, they prefer the alternative that is judged better by a majority of equally credible sources (Zhang, Hsee, &

Xiao, 2006). According to majority rule, a person should choose the Second City over the First City (in Figure 1) because two of three friends (*B* and *C*) gave higher ratings to that city.

Although Zhang, et al. (2006) concluded that people use majority rule for these decisions, they did not test its predicted violations of transitivity, which would provide clear evidence favoring it over transitive models. They did not test RBI, which would allow one to reject the entire class of models including Equation 1 and majority rule (Figure 4 and upper half of Table 1). Therefore, Experiments 1 and 2 investigate transitivity and RBI in this task to test majority rule and the more general class of models (Equation 1) that includes majority rule.

Methods of Experiments 1-2

Instructions read (in part) as follows: "Suppose you are trying to decide what city to visit on your vacation. Three friends give you advice by giving you their judgments of how much they think you will like each city. They made their ratings on a 1 to 10 scale in which 10 is the best and 1 is the worst."

Experiment 1

Each of the choice problems between cities were displayed using the format shown in Figure 1, where the numbers underneath *A*, *B*, and *C* represented ratings of the three friends. Participants clicked the button beside the city in each choice they would rather visit.

The first two trials were warm-ups, and the remaining trials provided two tests of RBI and two replications of a three-choice test of transitivity. The test of transitivity used six choices among the following three alternatives: $F = (4, 5, 5)$, $G = (2, 6, 7)$, and $H = (3, 7, 3)$. The three choice problems (*FG*, *GH*, and *HF*) were presented twice, with positions (First or Second City) counterbalanced; this meant that to be consistent, the participant had to press opposite buttons on choice problems *FG* and *GF,* which were separated by unrelated intervening choice problems and

presented in a random order. Note that majority rule predicts the intransitive cycle, $G \succ F$, $H \succ G$, yet $F \succ H$, which is denoted the *GHF* pattern. It is also useful to keep in mind that the mean ratings of *G*, *F*, and *H* are 5, 4.67, and 4.33, which yield the transitive order, $G \succ F \succ H$, denoted the *GGF* preference pattern. Both TAX and CPT predict *GGF* as well, when their "prior" parameters are applied to these values as if they were equally likely cash prizes (see Birnbaum, 2008a).

The first test of RBI used the following three choice problems: $S = (1, 4, 6)$ versus $R = (1, 2, 9)$, $S' = (5, 4, 6)$ versus $R' = (5, 2, 9)$, and $S'' = (10, 4, 6)$ versus $R'' = (10, 2, 9)$. The second test counterbalanced positions of the "safe" and "risky" cities, and used the following two choice problems: $S = (2, 4, 5)$ versus $R = (2, 2, 7)$ and $S'' = (9, 4, 5)$ versus $R'' = (9, 2, 7)$. RBI implies that $R \succ S$ iff $R' \succ S'$ iff $R'' \succ S''$, apart from random error.

There were two warm-up trials, followed by the intermixed experimental trials, presented in random order with the restriction that no two choice problems testing transitivity appeared on successive trials.

Participants were 214 undergraduates who served as one option toward an assignment in lower division psychology courses. Of these, 65% were female and 92% were 18-22 years of age. We tested participants until semester's end; we used this same stopping rule in Experiments 2, 3, and 4 with undergraduates from the same subject pool.

Experiment 2

Experiment 2 used a new sample of participants with a new experimental design that permitted tests of recycling as well as transitivity and restricted branch independence. In an attempt to find evidence of majority rule, the second experiment was constructed so that the average ratings

in the cities would be the same; perhaps some people who normally use mean ratings might use majority rule as a secondary strategy when the means are constant.

In the new design, $F = (4, 5, 6)$, $G = (5, 7, 3)$, $H = (9, 1, 5)$; $F' = (6, 5, 4)$, $G' = (5, 7, 3)$, and $H' = (1, 5, 9)$. In this design, $G' = G$, F' is a permutation of F and H' is a permutation of H. Note that majority rule in this case implies that $H \succ G$, $G \succ F$, but $F \succ H$; further, $F' \succ G'$, $G' \succ H'$, and $H' \succ F'$; this predicted pattern of intransitivity and recycling is denoted, HGF and $F'G'H'$. The average rating is 5 in all six alternatives. Each of these six choice problems was presented twice, with positions of the gambles counterbalanced and interspersed among other choices.

Experiment 2 also tested RBI, dominance, and included trials pitting majority rule against average value and majority rule against dominance under permutation. The first RBI design used $S = (1, 4, 6)$ versus $R = (1, 1, 9)$, $S' = (5, 4, 6)$ versus $R' = (5, 1, 9)$, and $S'' = (9, 4, 6)$ versus $R'' = (9, 1, 9)$; note that the average rating is the same within each choice problem. Each of these trials was repeated twice with positions counterbalanced. The second RBI design used $S = (1, 4, 5)$ versus $R = (1, 1, 9)$, and $S'' = (9, 4, 5)$ versus $R'' = (9, 1, 9)$; note that in this design the average of S is lower than that of R.

The following choice problem pitted majority rule against average value: $K = (5, 5, 5)$ versus $L = (6, 1, 6)$. The following pitted majority rule against dominance under permutation, $D = (2, 3, 5)$ versus $E = (6, 2, 4)$. Under majority rule, a person should choose D, but a permutation of E, $E' = (2, 4, 6)$ dominates D. These two trials were presented twice with position counterbalanced. In addition, there were 2 warm-up trials and 3 fillers (using transparent dominance) that were interspersed and presented in random order, with the restriction that no two trials testing transitivity or RBI would appear on successive trials.

There were 313 participants in Experiment 2; about half were recruited from the same pool as Experiment 1 and the rest were volunteers who were recruited via links on the WWW who participated via the Web at times and places of their own choosing. The undergraduate sample was recruited until the end of the semester and the Web sample until the end of the calendar year. Because the conclusions from these two subsamples were the same, they are combined in the following presentation. This sample was 67% Female; 49% were 18-22 years of age, and 12% were 40 years or older. Materials can be viewed at URL: http://psych.fullerton.edu/mbirnbaum/decisions/advice_friends3.htm

Results of Experiments 1 and 2

Experiment 1

The choice problems testing transitivity in Experiment 1 are analyzed in the first three rows of data in Table 2. The first row of the table shows that 60% chose $Y = G = (2, 6, 7)$ over $X = F = (4, 5, 5)$ in the first row, averaged over both replications. Similarly, 90% chose F over H (third row). These two modal choices agree with majority rule; however, only 17% chose H over G in the GH choice (second row), contrary to majority rule. In order to determine if the data underlying these binary proportions are consistent with a mixture of transitive orders, we next apply the TE model.

True and Error Model

Table 2 shows preference reversals between two presentations of the same choice problems to the same people (reversals are denoted XY^* or YX^* in Table 2), which are used to estimate error rates in the TE model. For example, the first row shows that 103 people chose G over F on both repetitions of this choice problem, but 25 + 24 = 49 people (23% of 214 participants) reversed preferences from one repetition to the other.

Insert Table 2 about here.

Let p represent the proportion of people who truly prefer Y in the XY choice problem, and let e represent the error rate for this choice problem, where $e < ½$, and errors are mutually independent. The probability of choosing gamble Y (and Y^*) on both presentations (of the XY and X^*Y^* choice problems, two repetitions of the same choice problem) is represented as follows:

$$P(YY^*) = p(1 - e)^2 + (1 - p)e^2 \qquad (6)$$

In other words, those people who truly prefer Y over X have correctly detected and reported their preference twice and those who truly prefer X have made two errors. The probability of switching from Y to X^* is given as follows:

$$P(YX^*) = p(1 - e)e + (1 - p)e(1 - e) = e(1 - e) \qquad (7)$$

This is the same as $P(XY^*)$, so the probability of reversals is $2e(1 - e)$. Finally, the probability of choosing X twice is $P(XX^*) = pe^2 + (1 - p)(1 - e)^2$. If descriptive, this model should reproduce the four observed frequencies corresponding to $P(YY^*)$, $P(YX^*)$, $P(XY^*)$, and $P(XX^*)$, which have 3 degrees of freedom (since the four sum to the number of participants). The model uses two parameters (p and e), leaving 1 df to test the model.

The estimated values of p and e for these three choice problems testing transitivity in Experiment 1 are given in Table 2. The $\chi^2(1)$ in the next column of Table 2 are tests of fit of this TE model. None are significant, showing that this TE model provides an acceptable fit. For comparison, the $\chi^2(1)$ testing response independence [i.e., $P(XX^*) = p(X)p(X^*)$, $P(XY^*) = p(X)p(Y^*)$, etc.], based on the same exact data entries, and using the same number of degrees of freedom, are also listed. These are all significant and all quite large, showing that we can reject any model that assumes response independence. These two tests highlight the distinction between error

independence and response independence, which has led to some confusion and unnecessary conflict in the literature (Birnbaum, 2013).

Table 3 shows the number of people who showed each pattern of preference responses on the first replicate, second replicate, and on both replicates in Experiment 1. For example, the last row of Table 3 shows that 11 displayed the intransitive, *GHF* pattern predicted by the majority rule on the first replicate, that 14 showed this same *GHF* pattern on the second replicate, and that only 2 of 214 participants showed this response pattern on both replicates (six choice problems). The most frequently observed response pattern, *GGF*, corresponds to the transitive response pattern matching the average ratings of the cities; 104 and 98 people showed this pattern on their first and second replicates, respectively, of whom 65 showed it perfectly on both replicates.

Insert Table 3 about here.

True and Error Model Analysis of Transitivity

The *g*TET model can be used to estimate the proportion of people who have each preference pattern. There are eight possible true preference patterns for the three choice problems, two of which are intransitive. Let p_{FGH}, p_{FGF}, ..., p_{GHF} denote the probabilities of these eight patterns. A person who obeys majority rule has the true preference pattern, *GHF*, so the estimate of p_{GHF} is an estimate of the proportion of participants who are consistent with majority rule, corrected for random errors.

Recall that each choice problem can have a different error rate, estimated from preference reversals, as in Table 2. The predicted probability of observing the intransitive pattern of majority rule, *GHF*, is the sum of eight terms for the eight true preference patterns, modified by errors. The predicted probability of showing the observed intransitive pattern on both replicates (all six trials) is therefore given as follows:

$$P(GHF, GHF) = \sum p(GHF, GHF \mid H_k)\, p(H_k) \qquad (8)$$

Where $H_1 = FGH$, $H_2 = FGF$, …, $H_8 = GHF$, are the 8 true patterns, and $p(GHF, GHF \mid H_k)$ are the conditional probabilities of showing each observed response pattern *GHF* on both replicates given each of the true patterns, and $p(H_k)$ are the probabilities of having each true pattern.

The conditional probability of showing response pattern, *GHF,* on both presentations of the same choice problems, given the person has the true pattern of $H_k = GHF$, is given as follows:

$$p(GHF, GHF \mid H_k = GHF) = (1 - e_1)^2(1 - e_2)^2(1 - e_3)^2, \qquad (9)$$

In this equation, we see that a person who has a true pattern of *GHF* has to avoid six errors in order to show the observed pattern. A person who had the opposite true pattern, *FGH,* would have to make six errors to show this opposite response pattern twice. Keep in mind that this model assumes that errors are mutually independent but this model does not imply response independence (Birnbaum, 2013). There are eight equations like Equation 8, each with 8 terms in the summation, for the probabilities of showing each observed pattern on both replicates. There are 64 equations for all possible response patterns.

We can partition the data into the frequency of showing each response pattern on both replicates and the frequency of showing each pattern on the first replicate but not both (Birnbaum, 2013), as in Table 3. This partition has sixteen mutually exclusive and exhaustive cells that sum to the number of participants (in this case, $n = 214$); therefore, there are 15 df in this partition of the data. For the parameters of the model, the true probabilities of the eight preference patterns will sum to 1, so there are 7 df to be estimated, and there are three error rates, leaving 5 df to test the model. This partition increases the cell frequencies while still providing constraint (degrees of

freedom) to estimate parameters and test the model.

The probabilities of the eight true patterns are estimated to minimize the $\chi^2(5)$ between predicted and obtained frequencies in Table 3. With error terms estimated from Table 2, the right-most column of Table 3 shows the best-fit estimates of the true probabilities for the 8 possible preference patterns. Only 1% of the participants were estimated to have the *GHF* pattern implied by majority rule (i.e., 2 people out of 214), whereas 60% are estimated to have the transitive pattern *GGF* as their true pattern. The test of TE fit to these 16 frequencies yields, $\chi^2(5) = 7.98$, which is an acceptable fit. As noted in Birnbaum (2013), one can also estimate both error rates and true probabilities simultaneously from Table 3 (instead of constraining the errors to fit only Table 2). With that method, the fit of *g*TET model is even better, $\chi^2(5) = 1.93$.

In order to test transitivity, we fit a special case of *g*TET with the assumption that no one was intransitive; i.e., two parameters were fixed to zero: $p_{FGH} = p_{GHF} = 0$, with error rates constrained to fit Table 2. The fit for this transitive model yielded $\chi^2(7) = 12.45$. The difference, $\chi^2(2) = 12.45 - 7.98 = 4.47$, is not significant, indicating that we cannot reject the hypothesis that everyone was transitive in Experiment 1.

To illustrate the power of these tests, 10 hypothetical participants who exhibited the predicted pattern of majority rule on both replicates were added to the data. In this case, the transitive model analysis yielded, $\chi^2(7) = 126.4$ (the critical value is 18.5 with $\alpha = 0.01$). Thus, if just 12 out of 224 had shown the predicted pattern of intransitivity on both repetitions (instead of only 2 out of 214), one could have confidently rejected the null hypothesis of transitivity.

We can reverse the above question and ask, given the actual data, how large a percentage

might actually satisfy majority rule by violating transitivity as predicted? Setting $p_{GHF} = 0.05$ and with all other true probabilities free, $\chi^2(6) = 12.04$, indicating that we could retain the hypothesis that as much as 5% of the sample might have satisfied majority rule (since 12.04 falls short of the critical value of $\chi^2(6) = 12.5$ with $\alpha = 0.05$); however, setting $p_{GHF} = 0.06$ or greater, we would reject this hypothesis with $\chi^2(6) = 13.44$ or greater. Therefore, one can retain the hypothesis that a few of our participants might satisfy majority rule, but one would reject the hypothesis that more than 6% of our sample did so.

This point deserves emphasis: Because majority rule has no free parameters, it must predict the intransitive pattern, *GHF*, apart from random error. Therefore, we can reject majority rule as a descriptive model of this study for the vast majority (at least 94%) of our participants.

This analysis should also address a concern that those unfamiliar with TE models might express. Someone might think that with so many parameters to estimate from the data, the TE model might not be able to distinguish cases where transitivity is satisfied from cases where it is not. As shown here, however, when the experiment includes replications of each choice problem in each block and has a sufficient number of participants, the TE model has enough power and precision to allow us to confidently reject transitivity if it had been violated by a small percentage of people and to place reasonably narrow statistical limits on the probabilities of transitive and intransitive response patterns.

Restricted Branch Independence

Tests of RBI are presented in Table 4 for Experiment 1. According to RBI, people should make the same responses in all the three choice problems, apart from error. According to RBI, if Source A gives both cities the same rating, it should not matter what the value of that common

rating was. However, when Source A rated both cities as 10, 66% chose the "riskier" alternative, $R'' = (10, 2, 9)$ over the safer one, $S'' = (10, 4, 6)$. However when Source A rated both cities as 1, then only 37% chose the "risky" alternative over the "safe" one [$R = (1, 2, 9)$ versus $S = (1, 4, 6)$]. Significantly more participants (81) switched from safe to risky as the common consequence was increased than switched in the opposite direction (19), $z = 6.2$, $p < .01$. By tests of correlated proportions, the differences between the first two rows in Table 4 and between the second and third rows are also both significant (49 to 21, $z = 3.35$; and 55 to 22, $z = 3.76$, respectively).

Insert Table 4 about here.

RBI also implies that people should make the same choice responses for the last two choice problems (last two rows) in Table 4. Instead, 19% prefer the risky (2, 2, 7) over (2, 4, 5) when Source A gave both cities a rating of 2; however, 48% preferred (9, 2, 7) over (9, 4, 5), when Source A gave both a rating of 9. In this case, 78 people made this preference reversal compared to 16 who had the opposite reversal ($z = 6.39$).

These significant violations of RBI refute not only models like Equation 1 (including majority rule) in which a constant feature should have no effect, they also violate the editing rule of "cancellation" and the constant-weight, averaging model of Equation 4 (sometimes called "anchoring and adjustment" or the "toting up" heuristic). They remain compatible with configural weight models (Equation 5) in which weights of the stimuli depend on their ranks (Birnbaum & Bahra, 2012a; Birnbaum & Zimmermann, 1998). However, they show the opposite trend from what is predicted from cumulative prospect theory's assumption that middle ranked values receive the least weight.

Although we infer from tests of RBI that people are not simply summing or averaging the ratings of the cities (which would imply RBI), it is also apparent in Table 3 that the most frequent

response pattern in the test of transitivity, *GGF*, corresponds to the order of the average ratings of the cities. This finding leads to the question: What if the means of the ratings were all equal, so people could not use this aspect of the cities to make their choices? When people cannot use mean ratings, perhaps they would use majority rule as a secondary strategy to decide. Therefore, Experiment 2 used cities with the same average ratings to search for evidence of intransitive cycling and recycling predicted by Equation 1 in this restricted domain.

Results of Experiment 2

The last six rows of Table 2 show choice percentages for the six choice problems used to test transitivity and recycling in Experiment 2. The estimates of true choice probabilities and error rates are also shown. In all cases, *g*TET fits the data better than the assumption of response independence, which is significantly violated in all six cases; one case showed a significant deviation from the *g*TET model.

Table 5 shows the analysis of transitivity and recycling in the two designs of Experiment 2, as in Table 3. The most common response pattern in both tables was the transitive order, *FGF*, which agrees with the order predicted by the TAX model applied to friend's ratings of cities as if they were dollar values of cash prizes in gambles, using prior parameters (Birnbaum, 2008a). However, unlike the results in Experiment 1, Table 5 shows that a small percentage of participants displayed response patterns predicted by majority rule.

Insert Table 5 about here.

When *g*TET is fit to the data (Table 5), using the estimates of error from Table 2, it is estimated that 18% of participants showed the intransitive *GHF* pattern and 15% showed the recycled intransitive pattern *F'G'H'*, consistent with majority rule. The opposite, intransitive patterns were rare (1% and 2%). The $\chi^2(5)$ representing the fit of *g*TET were 18.41 and 19.91 for

the two designs in Table 5, respectively (corresponding χ^2 values testing response independence were 727.5 and 1066.2, respectively). The special case TE models for the transitive case yielded, $\chi^2(7) = 289.3$ and 404.3, respectively, which are sharply higher. Thus, in Experiment 2, we have evidence that a significant minority of people (estimated 18% and 15% in the two designs) violated transitivity in the recycling manner predicted by majority rule, in a study where the mean ratings were kept constant. However, even in this experiment, the vast majority of participants were transitive.

Tests of Restricted Branch Independence

Because each choice problem testing RBI was replicated in the first RBI design of Experiment 2, we can analyze RBI using *g*TET. The first three rows of Table SM.1, in the On-line supplementary materials (SM) for this article, analyze the three choices of this design, as in Table 2. As Friend A's rating (the value on the common branch) improves from 1 to 9, the percentage choosing the risky gamble increased from 18% to 52%. In all cases, the tests of response independence were large and significant (smallest $\chi^2(1) = 90.8$) and *g*TET fit better than response independence; in one case, however, a test of *g*TET was significant ($\chi^2(1) = 20.6$).

Reference to Tables SM.1 and SM.2 occur around here.

Table SM.2 analyzes the response patterns in this test of RBI, as in Table 5. According to RBI, all of the response patterns should be *SS'S''* or *RR'R''*, apart from random error. However, many people switched from the safe to risky cities as the value on the common branch improved (the estimated true probabilities of response patterns *SS'R''* and *SR'R''* sum to 38% of the sample, whereas the sum of probabilities of all other violations of RBI was 3%). When the probabilities of all patterns violating RBI are set to zero, the index of fit jumps from $\chi^2(5) = 38.75$ to $\chi^2(11)$

=1551.0, a difference of $\chi^2(6) = 1521.25$. These are all significant.

Violations of RBI are compatible with Equation 5, which includes both the TAX model and CPT as special cases. According to the TAX model with prior parameters, the response pattern should have been *SS′S″*, which is the most common pattern observed; however, with different parameters, the TAX could also account for *SS′S″*, *SS′R″*, *SR′R″*, and *RR′R″*. The CPT model (with its prior parameters) predicts the response pattern, *RS′S″,* but with other parameters in an inverse-S weighting function, CPT could be compatible with *SS′S″, RS′S″*, *RR′S″,* and *RR′R″*. If CPT does not assume an inverse-S weighting function, however, it could account for other patterns, including those that are compatible with TAX (see Birnbaum, 2008a).

The second test of RBI in Experiment 2 showed that whereas only 27% chose (1, 1, 9) over (1, 4, 5), 59% chose (9, 1, 9) over (9, 4, 5). Of the 310 who completed both of these choice problems, 106 shifted from safe to risky as the common consequence was improved, compared with only 9 who reversed preferences in the opposite direction ($z = 9.05$), replicating the same type of violation of RBI.

Majority Rule Versus Expected Value and Permuted Dominance

According to majority rule, $L = (6, 1, 6)$ should be preferred to $K = (5, 5, 5)$ because two of three friends gave *L* the higher rating. Instead, 245 of 313 participants preferred *K* on both repetitions, compared to only 45 who chose *L* on both presentations. Table SM.1 shows that according to *g*TET, which fit this choice problem quite well [$\chi^2(1) = 0.04$], only 15% of the sample truly preferred *L*. We can reject the hypothesis that this true probability is greater than or equal to 0.21 ($\chi^2(1) \geq 5.45$), so this test indicates that we can reject majority rule for most people on this choice; we can also rule out the hypothesis that this true probability is less than or equal to 0.11,

$\chi^2(1) \geq 5.72$, so it appears that a certain percentage, likely between 11% and 21%, truly preferred *L*.

According to majority rule, *D* = (2, 3, 5) should be preferred to *E* = (6, 2, 4), even though *E* dominates *D* if the ratings can be permuted across the friends. Instead, 248 participants (81% of 310 who completed both presentations of this problem) chose *E* on both occasions compared to 27 (9%) who expressed the opposite preference on both repetitions. We can reject the hypothesis that more than 14% satisfied majority rule in this choice problem, $\chi^2(1) \geq 4.75$; we can also reject the hypothesis that 6% or fewer were compatible with it. For comparison, the permutation, *E′* = (2, 4, 6) was chosen over *D* = (2, 3, 5) by 97% of the sample.

Two other choice problems, (2, 2, 8) versus (2, 8, 8) and (4, 5, 9) versus (4, 4, 8) were compatible with transparent dominance for 99% and 96% of the sample, respectively. These cases appear compatible with perfect true conformity to transparent dominance, apart from random error.

Individual Participant Data

Examining individual data, 15 individuals were found to have data perfectly consistent with predictions of majority rule on all 12 choice problems testing transitivity and recycling. Because of the counterbalancing of stimulus presentations, participants had to use exactly opposite response buttons for all six choices with the same friend's ratings permuted over sources, and within each set of six problems testing transitivity, each response key had to be used exactly three times.

For comparison, there were 38 individuals whose data were perfectly consistent with the transitive pattern predicted by the TAX model with its prior parameters (*FGF*) for the same 12 choice problems.

For the 21 choice problems for which majority rule makes unambiguous predictions, the number of disagreements between predicted and observed response was calculated for each person's data. There were 31 people (10% of 313) who had 4 or fewer disagreements with majority

rule, of whom 5 were in perfect accord. However, among this group of 31, there were only 9 who chose (6, 1, 6) over (5, 5, 5) on both repetitions (only 33% of this group). So even within this selected group of participants whose responses best resemble predictions of majority rule, the rule does not continue to work when mean ratings of the cities are allowed to differ.

Thus, although most individuals satisfied transitivity in Experiment 2, and although the most common behavior showed perfect conformance to a transitive pattern that was invariant with respect to permutation, a small group of people showed evidence of the intransitive patterns of preference including recycling predicted by majority rule in the second experiment, when mean values were held constant in the choice problems.

Discussion of Experiments 1 and 2

The results of both studies allow us to reject majority rule as a description of how the majority of people made these choices based on advice. Anyone who followed majority rule should have been systematically intransitive; there is no flexibility of this theory in the designs used here, apart from random error. Allowing for error, the hypothesis that more than a small percentage was intransitive could be rejected in both studies. The hypothesis that no one was intransitive could be retained in Experiment 1 where the cities had different average values, but that hypothesis could be rejected in part of Experiment 2 (when cities in choice problems had the same mean value), in favor of the hypothesis that a few people were intransitive. Experiment 2 also showed that the recycling prediction failed for most participants but was observed in a few participants including a few who satisfied it perfectly.

Nevertheless, the data of both studies allow us to reject the theory that more than a small percentage of participants conformed to majority rule, even in the restricted domain where mean ratings were held constant. We conclude that majority rule is not an adequate descriptive theory of

choices, even as a secondary decision strategy in this restricted domain.

Although majority rule must be intransitive in these two experiments, the more general model of Equation 1 allows both transitive and intransitive response patterns, so Equation 1 cannot be rejected by a mere failure to find intransitivity. However, both studies found clear violation of RBI. Whereas it is possible that a person governed by Equation 1 might satisfy transitivity in these studies, it is not possible for such people to violate RBI.

The task used here is the same as that used by Zhang, et al. (2006), who concluded that their results indicated that people used majority rule. However, that study argued for majority rule without refuting transitive models, such as the constant weight averaging model (Equation 4). That study also did not apply any analysis of error.

To elaborate, the first experiment of Zhang, et al. (2006) consisted of a single choice problem, between City $X = (9, 6, 8)$ and $Y = (8, 9, 6)$. They found that 65% chose X over Y, which is consistent with majority rule. However, because they did not counterbalance the presentation order, their finding is also consistent with the rule: choose the first alternative. Their single choice problem also confounded serial position; note that if a person followed a constant weight averaging model in which there are weights for serial positions (a, b and c in Equation 4), who assigned lower weight to the middle branch (less weight for the middle source), that person could also prefer X to Y based on a weighted average. Consequently, this result is consistent with at least two simple rival models.

Zhang, et al. (2006) asked another group of participants to rate the attractiveness of these cities based on the friend's ratings. They found that 53% rated X higher than Y (perhaps some rated them equally). Zhang, et al. (2006) took this difference in percentage as evidence of preference reversals without considering that the error rates might differ for ratings and choices. They argued

that the change from 65% to 53% indicated that people were differentially using majority rule. Although they cited Lichtenstein and Slovic (1971), they did not apply the Lichtenstein and Slovic analysis of errors to distinguish real preference reversals from those attributable to error.

The "unpacking" manipulations used in Zhang, et al. (2006) are similar to what has been called "branch splitting" in studies of decision-making (Birnbaum & Navarrete, 1998). Such results are compatible with the transitive, configural weight model. When branches are split, they receive greater total weight than when they are coalesced. This means that splitting a higher evaluation will make an option seem better, but splitting a lower evaluation will make it seem worse. Thus, the results of the "unpacking" manipulation in Zhang, et al. (2006) are also compatible with transitive, configural weight models.

Therefore, their study provides no definitive evidence that anyone used majority rule or that anyone changed rules. The other studies of Zhang et al. (2006) also confounded serial position, presentation order, and error analysis; therefore, their conclusions regarding majority rule were not justified by their experiments.

In the case of majority rule, the model has no free parameters that allow it to predict transitive preferences in Experiments 1 and 2, so the results of both studies are stronger than a mere failure to find a possible result. Instead, the data refute majority rule for the vast majority of participants, because majority rule must violate transitivity in these studies, apart from error. Majority rule also failed in specific tests where the rule was pitted against average value and against dominance under permutation in Experiment 2.

In the tests of transitivity in Experiment 1, the average values of the three alternatives were not quite equal; they were 4.33, 4.67, and 5.00. With those stimuli, it was possible to reject majority rule for most people. The most frequent pattern observed was the transitive order

matching the averages. From these results, one might retreat to a position that a generalization of majority rule might still be descriptive, but only when the stimuli are constrained to have equal mean value. When the average values were equal in Experiment 2, a few showed intransitive cycles and recycles consistent with majority rule. Based on this finding, we constrained our next experiments to keep the average value equal in the choice problems testing transitivity, in an effort to use situations most favorable to this class of intransitive models. In addition, we investigate a task where it has been argued that people are systematically intransitive (Loomes, 2010): choices between dependent gambles, as in Figure 2.

Experiments 3 and 4: Testing Generalized Models

Regret theory was proposed as a theory of how people choose between gambles (Loomes & Sugden, 1982; Loomes, Starmer, & Sugden, 1991). It is interesting that although regret theory can also imply intransitive preferences, when it does so it makes opposite predictions from those of majority rule. As will be shown in this section, both regret theory and majority rule are special cases of a more general model that violates transitivity and exhibits recycling. This general model includes a transitive model, expected utility, as a special case.

Bernoulli developed expected utility (EU) theory (1738/1954) for risky decision making. EU is a special case of the transitive, constant weight averaging model (Equation 4) with $a = b = c = 1/3$. If the numbers of Red, White, and Blue marbles were independently manipulated (which is not done in our studies), then these weights should be equal to the probabilities of drawing Red, White, or Blue from the urn, according to EU.

To represent choices between dependent gambles (as in Figure 2), Equation 1 must be expanded slightly to allow for events that might have unequal objective or subjective probabilities.

A Family of Intransitive Models

A general model that includes both regret theory and majority rule as special cases is as follows:

$$A \succ B \text{ iff } \sum \phi(E_i)\psi(x_i, y_i) > 0 \qquad (10)$$

where $A = (x_1, E_1; x_2, E_2; \ldots; x_n, E_n)$ and $B = (y_1, E_1; y_2, E_2; \ldots; y_n, E_n)$ are two alternatives whose monetary consequences, x_i and y_i depend on the state of the world (*event*), E_i, where the n events are mutually exclusive and exhaustive ($i = 1$ to n); $A \succ B$ denotes that A is preferred to B. Note that we have slightly revised notation for gambles; whereas A, B, and C were advisors who rated cities in Experiments 1 and 2, A, B, and C (in Italics) represent gambles that yield cash prizes.

As in Equation 1, the contrast function, $\psi(x, y)$ satisfies the property that $\psi(x, y) = -\psi(y, x)$ hence $\psi(x, y) = 0$ if $x = y$ (c.f., Fishburn, 1982). Equation 10 extends Equation 1 by the additional terms, $\phi(E_i)$, which are the probabilities or *weights* of events E_i . [This term might be further restricted to be a subjective probability measure, by analogy with subjective expected utility (SEU; Savage, 1954); and when objective probability, p_i, is given, it might be further assumed that $\phi(E_i) = p_i$, as is done in Loomes et al. (1991).

Equation 10 includes regret theory (Loomes, et al., 1991), majority rule (Russo & Dosher, 1983; Zhang, et al, 2006), and most probable winner models (Blavatskyy, 2006) as special cases. With reference to Table 1, these theories can (and in some cases, must) be intransitive, they can (and sometimes must) show recycling, and they must satisfy RBI.

Majority rule (MR) can be represented as follows:

$$\psi(x, y) = \sigma(x, y) \qquad (11)$$

Where $\sigma(x, y) = 1$, 0, or -1 when $u(x) > u(y)$, $u(x) = u(y)$, or $u(x) < u(y)$, respectively, where $u(x)$ is

the subjective value or utility of consequence x. When probabilities are known and substituted for $\phi(E_i)$, this rule chooses the alternative with the highest total probability of yielding a higher outcome than the other alternative, summed across the events; for this reason it is also called the "most probable winner" model (Blavatskyy, 2006).

When the events are equally likely, as in Experiments 3-6 (where the numbers of Red, White, and Blue marbles of Figure 2 are equal), $\phi(E_i) = \phi(E_2) = \phi(E_3)$, in which case Equation 10 is equivalent to Equation 1; and with Equation 11, it reduces to majority rule. In Experiments 1 and 2, majority rule meant choosing the city that two of three sources rated higher, in Experiments 3-6, majority rule means choosing the gamble that yields a higher prize for two of the three mutually exclusive, equally likely events. Note that magnitudes of the differences have no effect in majority rule; only the signs matter. This limitation can be generalized, as is done below.

In regret theory, magnitudes of the contrasts are important. According to Loomes, et al (1991), regret is especially large in cases where the contrasts are greater. This property has been called *regret aversion* and is defined as follows: For any $x > y > z$,

$$\psi(x, z) > \psi(x, y) + \psi(y, z). \qquad (12a)$$

Although Expression 12a is defined as regret aversion, it could also be described as "rejoicing seeking," despite the different intuitive feelings aroused by these different emotional words.

The opposite relation (compared to Expression 12a) is a generalization of majority rule in which two small advantages tend to outweigh one large disadvantage spanning the same total gap in value:

$$\psi(x, z) < \psi(x, y) + \psi(y, z). \qquad (12b)$$

One might describe this opposite tendency as *advantage seeking* or as *disadvantage aversion*. This relation generalizes the basic idea of majority rule: whereas majority rule requires

that if two of three equally likely events lead to slightly better consequences, that alternative should always be chosen, no matter how small the two advantages and no matter how large the disadvantage under the third event, this more general version allows magnitude tradeoffs among the advantages and disadvantages.

A two-parameter, subtractive specification of the contrast function of Equation 10 that can represent regret aversion and advantage seeking (including majority rule and most probable winner) is the following:

$$\psi(x_i, y_i) = f[u(x_i) - u(y_i)] = \sigma(x_i, y_i)\left|(x_i)^{\alpha} - (y_i)^{\alpha}\right|^{\beta} \qquad (13)$$

where f is a strictly increasing function; $\sigma(x_i, y_i)$ is defined as in Equation 11; and α and β are constants. When $\beta > 1$, we have regret aversion (Expression 12a holds). When $\beta < 1$, however, we have advantage-seeking (Expression 12b). In the extreme, when $\beta = 0$, Equation 13 reduces to most probable winner, which is the same as majority rule when events are equally likely. When $\beta = 1$, we have a perfectly transitive model (Equation 4) that includes SEU, EU, and subjectively weighted utility (SWU) as a special cases, depending on whether $\phi(E_i)$ is assumed to be a subjective probability, objective probability, or probability weight, respectively. This model also assumes, $u(x) = x^{\alpha}$, where α is the exponent.

Although Expression 13 is quite flexible, it is not as general as Expression 10. This parameterized model is useful, however, because it can be fit to empirical choices between gambles, and the two estimated parameters show exactly where to search for violations of transitivity.

Expression 13 was fit to data by Bleichrodt, Cillo, and Diecidue (2010), who found that the median estimated β was slightly larger than 1.5 and median estimated $\alpha = 1$. Their data showed regret aversion in the aggregate and for the majority of individuals, analyzed separately (hence

$\beta > 1$). Experiments 3 and 4 will test for predicted intransitivity and recycling based on these estimates.

The logic of Experiments 3-4 is as follows: If systematic violations of transitivity can be found, they would refute the entire class of models that require transitive preferences (Table 1). According to regret theory, these estimated parameters inform us where to search for violations of transitivity and recycling, which if observed, refute the entire class of transitive models, including CPT, TAX, SWU, EU and others. Experiments 3 and 4 were designed to find those violations.

Regret aversion and advantage seeking could each lead to intransitivity, but the directions of the intransitive cycles are opposite for these two cases. For example, suppose there are three marbles in an urn: Red, White, and Blue. There are three prospects, defined as follows: A = (\$4, \$5, \$6), B = (\$5, \$7, \$3), and C = (\$9, \$1, \$5), where (x_1, x_2, x_3) are the consequences in a given prospect for drawing a Red, White, or Blue marble, respectively. To calculate predictions from Equation 13, let $u(x) = x$, with any $0 < \beta < 1$ for advantage seeking, and any $\beta > 1$ for regret aversion.

In the choice between A and B, advantage seeking (including majority rule) favors $B \succ A$, because B gives a higher payoff for two events, Red or White; hence, $B \succ A$. Similarly, $C \succ B$, because C has higher prizes for both Red and Blue. Finally, $A \succ C$ because A yields higher payoffs than C on both White and Blue; this completes the intransitive cycle.

Regret aversion ($\beta > 1$) also yields intransitive cycles, but it shows the opposite pattern of preferences (from those predicted by advantage seeking); namely, $A \succ B$, $B \succ C$, and $C \succ A$. This cycle is predicted because the largest differences between outcomes (which produce the greatest

regrets) cause the preferences: $A \succ B$, $B \succ C$, and $C \succ A$, as illustrated in Figure 5. In Figure 5, arrowheads indicate directions of preference; predictions for advantage seeking (majority rule) are shown in the outer circles with dashed lines, and predictions for regret aversion are shown in the inner circles. Predictions for the *ABC* design are shown on the left. Predictions for the *A′B′C′* design used to test recycling, are shown on the right and are described next.

Insert Figure 5 about here.

Recycling

According to Expression 10 (and Equation 13), it should be possible to find stimuli that violate transitivity, and such that the direction of the intransitive cycle can be reversed by suitable permutations of the same consequences over the events. The left side of Figure 5 shows that regret aversion implies clockwise intransitive cycle for *A*, *B*, and *C;* advantage seeking implies a counterclockwise cycle for these same stimuli. Let A' = (\$6, \$4, \$5), B' = (\$5, \$7, \$3), and C' = (\$1, \$5, \$9). In this case, $B' = B$, and A' and C' are permutations of *A* and *C*, respectively. By permuting the same consequences, we should observe a clockwise cycle (right side of Figure 5) under advantage seeking (majority rule) and counter clockwise cycle under regret aversion. This prediction of reversible cycles (*recycling*) is tested in Experiments 3 and 4.

Figure 6 shows the predicted preference response patterns from Equation 13 when both α and β are free to vary. Predictions are for *AB, BC* and *CA* choices of Figure 5 and for their permutations, *A′B′, B′C′* and *C′A′* respectively. Predictions are based on the values in Figure 5 multiplied by 10, ranging from \$10 to \$90, as used in Experiment 4, and they assume that events are equally likely. The code, *1,* denotes preference for alphabetically higher item, no matter what order the stimuli were presented; e.g., *1* indicates preference for *A* in either the *AB* or *BA* choice. The recycling pattern predicted by regret aversion ($A \succ B$, $B \succ C$, yet $C \succ A$ but $A' \prec B'$, $B' \prec C'$, yet

$C' \prec A'$) is denoted, *111 222*. Note that the regions in which both intransitivity and recycling are predicted (patterns, *111 222* and *222 111*) are quite large, but do not fill the entire space. There are also regions where one design shows intransitivity and the other does not.

The estimated values from Bleichrodt, et al. (2010) for regret theory ($\alpha = 1$ and $\beta = 1.5$) are well inside the region of predicted intransitivity and recycling (pattern *111 222*) for both Experiment 4 (shown in Figure 6, prizes from \$10 to \$90) and Experiment 3 (with prizes from \$100 to \$900; which is similar to Figure 6). If some of our participants use parameters similar to those in previously published studies, those participants should show both intransitivity and recycling in Experiments 3 and 4.

Insert Figure 6 about here.

True and Error Model

In Experiments 3-4, we again use a TE model to represent the variability of responses to choice problems. As in Experiments 1-2, error rates are estimated from preference reversals when the same choice problem is presented to the same participant twice in the same block of trials. This model is expanded to include both intransitivity and recycling in Appendix A.

Previous Tests of Transitivity of Preference in Gambles

Tversky (1969) reported violations of WST in preferences among binary gambles predicted by a lexicographic semiorder model and concluded that about a third of the people tested were systematically intransitive. However, the methods of analysis used by Tversky have been criticized, and recent studies failed to find convincing evidence for the predicted intransitivity (Birnbaum & Bahra, 2012b; Birnbaum & Gutierrez, 2007; Regenwetter, et al., 2010, 2011).

Not only have recent studies failed to find much, if any, evidence for the intransitive preferences predicted by lexicographic semiorder models, including the priority heuristic, but other

research found that a majority of people tested systematically violated integrative independence, interactive independence, and priority dominance-- three critical properties implied by lexicographic semiorder models (Birnbaum & LaCroix, 2008; Birnbaum, 2010). This class of models, depicted in the lower right of Table 1, has been refuted by systematic violations of its critical properties.

Equation 10 and its special cases (Equation 13) do not imply the same kind of intransitivity as do lexicographic semiorders, nor does Equation 10 imply integrative independence, interactive independence or priority dominance. Therefore, failures to find evidence of intransitivity in the paradigm used by Tversky (1969) and empirical violation of theorems implied by lexicographic semiorders (Birnbaum, 2010) do not address the viability of Equation 10, including regret theory and majority rule for choices between gambles.

Although initial reports contended that people violate transitivity where predicted to do so by regret aversion (Loomes, et al., 1991), alternative explanations compatible with rival models have been proposed for those results (Birnbaum & Schmidt, 2008; Humphrey, 2001; Starmer & Sugden, 1993; Sopher & Gigliotti, 1993; Leland, 1998). For example, it was argued that if one type of intransitive cycle was more frequent than the other, it might be evidence that people are intransitive; however, asymmetry of incidence of different types of intransitive cycles does not rule out transitive models, if it is possible that responses contain error; nor would equality of different types of intransitive patterns rule out intransitive models (Birnbaum & Schmidt, 2008).

The perceived relative arguments model (PRAM) by Loomes (2010) can violate transitivity. However, the evidence cited by Loomes (2010) relied on asymmetry of different types of intransitive cycles; the experiments he cited did not analyze data with a proper error model. The TE model can easily predict asymmetry of incidence of different intransitive patterns in cases where no

one is ever intransitive. Therefore, it remains an open question whether people actually show the intransitive behavior predicted by regret theory or PRAM.

Failures to find evidence of intransitivity and possible theoretical reinterpretations of dubious evidence do not rule out the possibility that regret aversion or advantage seeking might produce real violations in a properly devised and analyzed study. Experiments 3 and 4 correct this deficiency in the literature.

Restricted Branch Independence

For three-branch gambles such as described above, RBI can be written as in Expression 2, except x, y, and z now represent consequences to be obtained if a Red, White, or Blue marble is drawn, instead of ratings of cities. RBI should hold for choices among gambles if the consequence is the same under any of the events (no matter the value of that constant consequence), and this implication holds whether events are equally likely or have any other set of fixed probabilities. RBI also requires that the number of branches is the same in all four gambles of Expression 2.

According to Equation 10, any event-consequence branch that is the same in both alternatives contributes nothing to the preference between the alternatives, because $\phi(E_i)\psi(z, z) = \phi(E_i)\psi(z', z') = 0$, for all such common (constant) consequences. Therefore Equation 10 joins Equation 1 in the upper right corner of Table 1 since it violates transitivity, exhibits recycling, and satisfies RBI.

Methods of Experiments 3 and 4

Experiment 3

In Experiment 3, participants viewed the materials on computers in the lab and made choices between gambles. Each person responded to each of 30 choice problems (twice), clicking a button beside the gamble in each pair that they would rather play. Gambles were described in terms

of an urn containing 99 otherwise identical marbles that differed in color: 33 were Red, 33 White and 33 Blue. A marble would be drawn at random and the color of marble would determine the prize, according to the payoff matrix presented for each choice problem. Each trial was displayed as in Figure 2, except without coloring of the columns shown in Figure 2; instead, each table had a uniform color.

Participants clicked a button beside the First or Second Gamble to indicate their preferences. Each of the choice problems from the main designs testing transitivity, recycling, and RBI was presented twice (separated by intervening trials) in a block of trials with the alternative gambles counterbalanced by position (First or Second). This meant that for a person to be consistent, she or he had to appropriately switch response buttons in order to designate the same preference response.

Table 6 shows the choice problems used in the transitivity and recycling designs. Note that the gambles *A, B, C, A′, B′,* and *C′* are simply 100 times the consequences illustrated in Figure 5. Values in Experiment 4 were only 10 times those in Figure 5.

Insert Table 6 about here.

There are 6 choice problems in the *ABC* and *A′B′C′* designs testing transitivity and recycling; with the counterbalanced repetitions, there are 12 trials in these designs. These were separated from each other by at least one trial from filler or RBI designs. In addition, each person completed each block of 30 trials twice, separated by intervening tasks that required about 10 minutes. That means that each choice problem of these main designs was presented four times to each person, twice in each of two blocks of trials.

There were two RBI designs. The first used S = (\$100, \$400, \$600) versus R = (\$100, \$100, \$900), S' = (\$500, \$400, \$600) versus R' = (\$500, \$100, \$900) and S'' = (\$900, \$400, \$600) versus R'' = (\$900, \$100, \$900); the second RBI design used S = (\$100, \$400, \$500) versus R = (\$100,

$100, $900) and S' = ($900, $400, $500) versus R' = ($900, $100, $900). In addition, two trials pitted majority rule against expected value: ($500, $500, $500) versus ($600, $100, $600) and ($200, $300, $500) versus ($600, $200, $400). There were also three trials testing transparent dominance.

Complete instructions and trials can be viewed from the following URL: http://psych.fullerton.edu/mbirnbaum/SPR_07/choice_regret.htm

Participants were 240 undergraduates enrolled in lower division psychology courses, 69% were female; 97% were 22 years of age or younger (73% were 18 years) and none were over 40.

Experiment 4

Because a few participants in Experiment 3 showed evidence of intransitivity and recycling, Experiment 4 was designed in an attempt to increase the incidence of intransitive behavior; in addition, a "control" design was added. Experiment 4 used a new sample of participants with similar stimuli and general procedures, except for the following modifications: We modified the stimuli by coloring the backgrounds of columns in Figure 2 with their appropriate colors. We thought that coloring by events might increase comparisons within columns (events) and thus increase the incidence of intransitive and recycling behavior. The consequences in the tests of transitivity in Experiment 4 were one-tenth of those used in Experiment 3, ranging from $10 to $90.

There were 42 choice problems, including all choice problems from Experiment 3 plus a new, $A''B''C''$ control design using permutations of A, B, and C, in which consequences were put in ascending order for Red, White, and Blue, respectively; i.e., A'' = ($40, $50, $60), B'' = ($30, $50, $70), and C'' = ($10, $50, $90). This design is a "control" because these stimuli should not produce intransitive cycles, according Equation 10. In addition, the filler designs were expanded so that any two trials testing transitivity were always separated by at least one intervening trial.

Materials for Experiment 4 can be viewed at the following URL: http://psych.fullerton.edu/mbirnbaum/SPR08/choice_trans_02_05_08.htm

The participants in Experiment 4 came from two sources: 116 undergraduates from the same subject pool as in Experiment 3 were tested in the lab (until the end of the semester); in addition, 289 volunteers participated via the Internet, making a total of 405 participants. Although we stopped recruiting at the end of the semester, Web volunteers continued to participate. We used all data collected by the time we were ready to analyze the results. Of the entire sample, 61% were female, and 49% were older than 22 years, including 12% older than 40 years.

In both Experiments 3 and 4, prizes were hypothetical; participants were told, "Imagine that when this study is over, you will play one of your chosen gambles for real money." Real cash incentives were employed in Experiments 5 and 6.

Consistency of the Choice Responses

In Experiment 3, there were 11 choice problems that were repeated within each block with positions of the alternatives counterbalanced, and all 30 trials were presented in two blocks. The mean percentage agreements within-blocks were 75% and 79% in the first and second blocks, and the agreement between-blocks over all 30 choice problems was 78%. In Experiment 4, there were 16 choice problems repeated within block (and there was just one block). The mean within-block consistency was 82% for these 16 choice problems. Because positions of the alternatives were perfectly counterbalanced within blocks, if a person repeatedly pressed the same button, that person would show 0% agreement within blocks.

Results of Experiments 3 and 4

Tests of Transitivity and Recycling

Table 7 shows frequencies of response patterns in the *ABC* design of Experiment 3 for

counterbalanced presentations, combined over blocks. Response patterns are designated by Italics, where *1* and *2* indicate preference for the gamble with higher or lower alphabetic label in a choice problem, respectively. For example, the response pattern, *112*, represents observed preference for *A*, *B*, and *A* in the *AB*, *BC*, and *CA* choices, respectively; i.e., $A \succ B$, $B \succ C$, and $A \succ C$, respectively.

Entries on the major diagonal indicate the number of cases where participants made the same responses in all three choice problems in both repetitions, independent of positions. We use the terms "repeated" or "consistent" to refer to these cases.

Insert Table 7 about here.

The most frequently repeated pattern of responses was *112*. The entry of 101 on the diagonal in the table indicates that the *112* pattern was observed in both replicates of a trial block 101 times. This response pattern is predicted by the transitive, TAX model (Equation 5) with "prior" parameters [parameters that had been used in advance to predict results in a number of studies (Birnbaum, 2008a, 2010): $u(x) = x$, $a = b = c = 1$, $w_L = ½$, $w_M = 1/3$, and $w_H = 1/6$]. The next most frequently repeated response patterns are *212,* and *221*, which are also transitive. Recall that EU is a special case of Expression 10 (it is a special case of both TAX and CPT), and EU is also a special case of Equation 13 when $\beta = 1$. Note in Figure 6 that when $\beta = 1$, Equation 13 implies response pattern *112* when $\alpha < 1$ and *221* when $\alpha > 1$.

The pattern *111* is the intransitive cycle predicted by regret theory for the *ABC* design, based on parameters observed by Bleichrodt, et al. (2010): $\beta = 1.5$ and $\alpha = 1$, or with parameters near those values (Figure 6). The pattern *222* is the intransitive cycle predicted by advantage seeking (generalized majority rule), also illustrated in Figure 6. The intransitive cycle predicted by advantage seeking (lower right of Table 7) was repeated (on both repetitions) 18 times, and the

pattern predicted by regret aversion was repeated only 2 times (upper left corner of Table 7).

Table 8, arranged as in Table 7, shows the corresponding frequencies of response patterns when the consequences were permuted, in the *A'B'C'* design. Again, the same transitive patterns *112*, *212*, and *221* are the most frequently repeated patterns (diagonal of Table 8). Comparing Tables 7 and 8, note that the most frequently repeated intransitive pattern in Table 8 is *111* (upper left corner of Table 8) instead of *222* (lower right corner of Table 7). In this design, advantage seeking predicts *111* and regret aversion predicts *222*, and the pattern of advantage seeking (generalized majority rule) was again more common than that of regret aversion.

Insert Table 8 about here.

Table 9 analyzes recycling patterns. Rows in Table 9 represent the response patterns in the *ABC* design and the columns represent response patterns in the *A'B'C'* design. The most common response pattern is *112 112* (i.e., the *112* pattern in both permutations of the same consequences, *ABC* and *A'B'C'*), followed by *212 112*, *112 212*, *221 221*, and *212 212*. The recycling response pattern of advantage seeking (in the lower left of Table 9) corresponding to *222 111* is more frequent (27 instances) than the opposite pattern predicted by regret aversion, *111 222* (9 instances).

Insert Table 9.

These main results of Experiment 3 were replicated in Experiment 4; details are presented in the Online supplement, in Tables SM.3, SM.4, and SM.5, which correspond to Tables 7, 8, and 9 from Experiment 3. Again, the most frequent response pattern in all three tables was *112 112* (transitive). The more frequently repeated intransitive pattern in the *ABC* design (Table SM.3) was again *222*, as in Table 7 (10 instances of pattern *222* out of 405 compared to 1 instance of the *111* pattern), and in the *A'B'C'* the more common intransitive pattern was *111*, as in Table 8 (11 instances of pattern *111* compared to 0 instances of *222*). Finally, in Table SM.5 (corresponding to

Table 9) the more frequent intransitive recycling pattern was *222 111* pattern (18 instances) rather than the *111 222* pattern (4 instances), as in Table 9. As in Experiment 3, transitive patterns were the most frequent, especially the pattern *112*, which was repeated in 116, 161, and 261 instances in Tables SM.3, SM.4, and SM.5, respectively.

Reference to Tables SM.3, SM.4, and SM.5.

In the *A"B"C",* "control" design of Experiment 4, the consequences were listed in ascending, ranked order for Red, White and Blue events, and so Equation 10 no longer predicts intransitivity. In this control design, the fewest repeated, intransitive response patterns were found, consistent with the theory that the permutation of consequences over events was the key to the systematic, if infrequent intransitive patterns observed. Detailed results are presented in Table SM.6, which showed only 1 and 3 instances of the repeated response patterns *111* and *222*, respectively, compared to 149 instances of the repeated pattern *112*.

Reference to Table SM.6.

True and Error Model Analysis of Transitivity and Recycling

We fit TE models to the data in Tables 7, 8, and 9 of Experiment 3 and to the data of Tables SM.3, SM.4, SM.5, and SM.6 of Experiment 4. These models, described in Appendix A, are used to estimate the relative incidence of transitive and intransitive response patterns, accounting for response errors. The estimated parameters are presented in Table 10. In Experiment 3, the index of fit (ΣG^2) is the sum of three, redundant G^2, each of which is defined on one table of Experiment 3. The G^2 is an index of fit that is similar to the conventional Chi-Square and is regarded as a better measure when frequencies are small; the larger the index, the worse the fit (Appendix A).

In Experiment 4, the index is the sum of four G^2 indices, each computed on one of the four tables in that study. The procedure of adding these indices across tables means that the models are

constrained to fit all tables with the same parameters, but it also means that these indices of fit are inflated by addition of redundant information (responses by the same people are re-counted in different ways in these different tables). Therefore, they cannot be interpreted for statistical significance; however, we do think that they provide descriptive metrics for comparing different variations of the models. Those parameters that were fixed or constrained are shown in parentheses.

Insert Table 10.

The general TE model of Table 10 allows transitive and intransitive preference patterns, as well as both possible recycling patterns (*111 222* and *222 111,* labeled R and MR, respectively). The first model (first column) allows six different error terms for choice problems *AB, BC, CA*, *A′B′, B′C′* and *C′A′* in Experiment 3, and it allows three additional error rates in Experiment 4 for the three choices in the *A″B″C″* design as well (f_1, f_2, and f_3). However, as shown in the second column for each experiment, constraining the error terms to be independent of permutations (e.g., the same for *AB*, *A′B′*, and *A″B″*) did not reduce the fit by more than a tiny amount, nor did the assumptions that three of the terms representing intransitivity (including regret aversion) are zero: $p_{111} = p_{222} = p_R = 0$.

The models labeled MR in Table 10 are special cases of the general TE models: these allow all transitive patterns, plus one parameter (p_{MR}) in each study representing the proportion of people who are estimated to be following the intransitive and recycling predictions of advantage seeking (generalized majority rule, MR). The model labeled "Transitive" is a special case of this model in which this last source of intransitivity and recycling is set to zero ($p_{MR} = 0$). Notice that this one parameter produced larger changes in fit in both Experiment 3 and Experiment 4 than those due to the six (or nine) parameters representing additional error rates and other intransitivity, including regret aversion ($p_R = 0$) combined; that is, the difference in fit between general model and MR is

less than that between MR and Transitive models in both experiments.

The sum of the estimated probabilities of transitive response patterns exceeds 90% in all variations of the TE model in both studies. Note that the transitive models were also assumed to be independent of permutation. Even with this restriction, the vast majority of participants (more than 90%) produced data that can be represented by such transitive models. The most frequent response patterns are transitive patterns, *112*, *212*, and *221*, with more than half of the sample estimated to conform to the pattern, *112*, predicted by the prior TAX model in all models fits of both Experiments 3 and 4.

The estimated incidence of intransitivity and recycling consistent with advantage seeking (majority rule), though small (8% and 6% of the participants in Experiments 3 and 4, respectively), appears to be real. Comparing the predicted frequencies against the observed ones, we think the change in fit due to the one parameter, p_{MR}, is large enough to represent a true effect for a small proportion of the sample. With this term, the model was able to account for small, but systematic trends apparent in Tables 7-9 and SM.3-SM.6. The estimated incidence of intransitivity and recycling in Experiment 4 is lower than that in Experiment 3, providing no support for our hunch that coloring the backgrounds of the columns (or the other variations of procedure) would increase the incidence of intransitive and recycling behavior.

When we fit the corresponding model for regret aversion, with p_R free and with p_{MR} set to zero, the indices of fit in Experiments 3 and 4 were 247.7 and 526.7 (compared to 248.4 and 527.6 for the fully transitive models in Table 10), respectively, and estimated p_R for regret aversion were less than 0.01 in both studies. From these analyses and comparing predictions of models against the data in the tables, we conclude that no convincing evidence was found to refute the hypothesis that no one exhibited regret aversion. That is, data fit the hypothesis that no one did what they were

predicted to do according to regret theory, given the parameters of regret theory estimated by Bleichrodt, et al. (2010).

Examining complete response sequences for all 24 choice problems testing transitivity in Experiment 3, we found two people who were perfectly consistent with the predictions of intransitivity and recycling for advantage seeking (majority rule) on all 24 choice problems (2 blocks by 2 repetitions per block by six choices, *AB*, *BC*, *CA*, *A'B'*, *B'C'*, and *C'A'*). In addition, there were 3 others who were perfectly consistent in one block (12 choices) but not the other. In Experiment 4, there were 7 who were perfectly consistent with MR on all 12 responses (1 block by 2 repetitions by 6 choice problems). No one was perfectly consistent with the predictions of regret theory in either study.

Although the relative frequency of people who showed evidence of intransitive cycles and recycles is small, analysis of complete response patterns strengthens the arguments that it is real. For example, the finding that two people showed perfect consistency with majority rule in 24 out of 24 choices, including perfectly opposite responses when the same alternatives were presented in opposite order or when consequences are merely permuted, seems too unlikely to occur by chance.

The TE models use estimated error rates to estimate the incidence of intransitivity (Table 10); however, one can also compute the probability of such an occurrence (24 out of 24 responses perfectly intransitive and recycling) using simpler, standard binomial calculations, with the conservative assumption that error rates are less than ½ (otherwise we call them "lies" rather than "errors"). We have been unable to devise a single transitive theory or mixture of theories in which such a person has not made at least 12 errors. If the probability of each error is less than ½, the probability of 24 out of 24 responses consistent with majority rule is less than $(1/2)^{12} = 0.000244$; however, the opposite pattern of results (regret aversion) would also be noteworthy, so the

probability to show either combination of intransitivity and recycling (i.e., *111 222* or *222 111*, is 0.000488. The binomial probability to observe 2 or more such cases out of 240 independent participants in Experiment 3 is $p = 0.0064 < .01$. If someone can construct a more probable null hypothesis account of 24 responses out of 24 tests (perfectly intransitive, with perfect recycling, and perfect consistency with counterbalancing), please let us know.

It is interesting and perhaps ironic that these few cases showing intransitive behavior are opposite the predictions of regret theory that was developed to account for violations of expected utility theory; it means we cannot use Equation 10 to account for these intransitive results and also use Equation 10 to describe the common types of violations of EU.

Although a few showed cycling and recycling, the most common response patterns were transitive and invariant with respect to permutation. There were 21 participants in Experiment 3 who were perfectly consistent with the transitive pattern, *112*, on all 24 choice problems and 35 who were perfectly consistent in one block but not the other. In Experiment 4, there were 83 who were perfectly consistent with the transitive pattern *112* on all 12 problems, including 66 who also showed the same pattern perfectly on the *A″B″C″* choice problems that were included only in Experiment 4.

Tests of Restricted Branch Independence

Table 11 includes the choice proportions from both studies for the tests of RBI and certain other choice problems. Note that the choice proportions (choosing the riskier gamble) in the first three rows increase from 0.17 to 0.48 in Experiment 3 and from 0.16 to 0.40 in Experiment 4 as the value of the common consequence is increased.

These results are similar to those of Experiments 1 and 2 and Birnbaum and Bahra (2012a), who found systematic violations of RBI in individuals. These violations are opposite in direction

from the predictions of CPT with an inverse-S shaped weighting function, but are consistent with the type of violation implied by the TAX model and reported in previous studies (Birnbaum, 2008a). The fourth and fifth rows of Table 11 show a similar result in the second RBI design.

Insert Table 11.

Violations of RBI are not consistent with the Expression 10. Therefore, while a small percentage might show the pattern of violations of transitivity and recycling implied by advantage seeking, our data also refute Equation 10 as a general description for all participants.

The next two rows (MR-EV) in Table 11 show that most participants in both studies (more than 80%) preferred gambles with higher expected value rather than those favored by majority rule, when these two features were pitted against one another. The last three rows show that the vast majority of choice responses were consistent with transparent dominance in both studies.

Discussion of Experiments 3 and 4

Most participants exhibited transitive preferences, but a few people showed the pattern of intransitivity and recycling predicted by Equation 10 under advantage seeking (generalized majority rule or most probable winner model) and not under regret aversion.

When a small proportion of the participants show a certain effect, such as the estimated 8% in Experiment 3 who showed evidence of intransitivity and recycling, it seems reasonable to ask if there might be some experimental manipulation that might make that percentage higher or lower. We tried to increase the incidence of intransitivity in Experiment 4 by color-coding the columns of the display (Figure 2), and by recruiting a more highly educated sample (who we thought might be more motivated and show lower rates of error). Although these new participants showed higher within-block consistency than those in Experiment 3, we found no evidence of any higher incidence of intransitive behavior in Experiment 4 than in Experiment 3 (in fact, the estimated incidence was

slightly lower). Recall it was estimated that 8% and 6% of the sample followed the predictions of advantage-seeking in Experiments 3 and 4, and that no one evidenced regret aversion. In Experiments 5 and 6, we attempt more desperate measures in our search for intransitive behavior.

Experiments 5 and 6: Desperately Seeking Intransitivity

The experimental designs of Experiments 3 and 4 were chosen to detect violations where predicted by the parameters of regret theory, as fit to previous empirical data. They detected no evidence of anyone following regret theory, and the infrequent but systematic violations of transitivity that were observed were of the opposite type from those predicted by previously estimated parameters. However, those experiments left a way out for Equation 10 to be excused for its failure to predict the results. This way out is illustrated in the transitive regions of Figure 6; perhaps no one in our study had parameters close to those reported by Bleichrodt, et al. (2010). Furthermore, a defender of Equation 10 might argue that Figure 6 is based on Equation 13, which is a special case of Equation 10. Perhaps some other functions might allow Equation 10 to slip out of predicting intransitivity in Experiments 3 and 4.

Experiments 5 and 6, however, make use of an experimental design in which even the most general form of Equation 10 must imply intransitivity and recycling if people are either regret averse (Expression 12a) or advantage seeking (Expression 12b).

A New Design Testing Transitivity

In this design, let $I = (x, y, z)$, $J = (z, x, y)$, and $K = (y, z, x)$, where $x > y > z > 0$. If there is regret aversion (Expression 12a), Equation 10 implies intransitive choices of the form, $I \succ J$, $J \succ K$ but $K \succ I$. Proof: By regret aversion, $\psi(x, z) > \psi(x, y) + \psi(y, z)$ for all $x > y > z$. From Equation 10, if the three events are equally likely [$\phi(E_i) = \phi(E_2) = \phi(E_3)$], it follows directly that $I \succ J$, $J \succ$ K and

$K \succ I$. If we assume advantage seeking (Expression 12b), $\psi(x, z) < \psi(x, y) + \psi(y, z)$, we see that Equation 10 implies the opposite intransitive pattern, which can be proved in the same way. Therefore, if people are either regret averse or advantage seeking, they must be intransitive in this design, and if they are regret neutral, $\psi(x, z) = \psi(x, y) + \psi(y, z)$, they must conform to EU and be indifferent in all three choices.

Furthermore, we can construct permutations that must show recycling: Let $I' = (z, y, x)$, $J' = (x, z, y)$, and $K' = (y, x, z)$; these permutated gambles produce opposite preferences, $I' \prec J'$, $J' \prec K'$ and $K' \prec I'$, according to regret aversion and the opposite intransitive cycle according to advantage seeking. Therefore, this design forces this family of models (Equation 10) to show the complete pattern of cycles and recycles illustrated in Figure 5, as long as they are not always perfectly indifferent.

With reference to Equation 13 (special case of Equation 10) and Figure 6, the entire space should produce intransitive cycles in this design, except where $\beta = 1$, where a person is always perfectly transitive. When $\beta > 1$, we should see "regret" cycles (*111* in *IJK* design), and when $\beta < 1$, we should observe "advantage-seeking" cycles (*222* in *IJK*). In other words, if people are not always transitive ($\beta = 1$), then they *must* be intransitive in this design.

The use of this experimental design can be viewed as an act of desperation because in a sense, these choice problems are degenerate from the viewpoint of transitive theories (such as Equations 4 and 5), if equal weights are assigned to the three equally likely outcomes of Red, White, and Blue. Would you rather have I = ($90, $50, $10) or J = ($10, $90, $50)? If a person considers these to be equally attractive, she might either choose arbitrarily (perhaps randomly) or seek some tiebreaking rule to decide. Perhaps that tiebreaker would be regret aversion or advantage

seeking.

If anyone were inclined to use regret aversion or attraction seeking (including majority rule)—even if only as a secondary, tie-breaking rule in this limited situation--this experiment gives her or him a perfect opportunity to do it. A regret averse person would choose I = ($90, $50, $10) over J = ($10, $90, $50) to avoid the regret under Red of getting only $10 rather than $90 (which exceeds the sum of the other two regrets), and the advantage seeker (including majority rule) would choose J because under two colors (White and Blue), she would get a better prize with J, and the sum of these two advantages exceeds the disadvantage under Red.

On the other hand, if people do not show intransitivity and recycling in this design, then we can reject Expressions 10 (the most general form of regret and majority rule), except for its degenerate case where it implies EU and where people are *always* transitive. This experimental design is an extremely friendly environment for Expression 10 because many rival theories allow no way to choose among these alternatives, whereas Expression 10 provides clear preferences. So if we do not find the predicted intransitivity in this "desperate" design, we should not expect to find it anywhere else.

PRAM and Salience Weighted Utility Models

This new experimental design also tests the perceived relative argument model (PRAM) of Loomes (2010) and the salience weighted utility model of Leland and Schneider (2014). According to PRAM, people compare the relative arguments for each alternative in a choice problem and choose the alternative with the stronger arguments. This model is represented as follows:

$$J \succ I \text{ iff } \pi(b_J, b_I) > \psi(y_I, y_J) \qquad (14)$$

Where $\pi(b_J, b_I)$ is the perceived relative advantage of J over I based on probabilities, and $\psi(y_I, y_J)$ is the perceived advantage of I over J based on the consequences. It is assumed that b_J and b_I are

the probabilities that J gives a higher consequence than I, and vice versa, respectively (Loomes, 2010). It is assumed that $\psi(y_I, y_J)$ depends on contrasts between consequences of I and J for the same events (Loomes, 2010, Footnote 19). Although Loomes (2010) has further specified these functions in parametric forms, we can test this general form of Expression 14 using the new *IJK* design of Experiments 5 and 6 without these additional assumptions.

In this design, $I = (x, y, z)$, $J = (z, x, y)$, and $K = (y, z, x)$, where $x > y > z > 0$. Note that $\pi(b_J, b_I)$, the relative argument of probability, is the same in all three choices: J is better than I for two of three equally likely colors, K is better than J for two of three, and I is better than K for two of three. Therefore, $\pi(b_J, b_I) = \pi(b_J, b_K) = \pi(b_K, b_I)$. Similarly, the relative argument based on consequences is also the same in all three choices; i.e., I is better than J (J is better than K, and K is better than I) because it has x instead of z, whereas J is better than I (K is better than J, and I is better than K) because it has x instead of y and y instead of z, respectively. Therefore, $\psi(y_I, y_J) = \psi(y_J, y_K) = \psi(y_K, y_I)$. It follows from Equation 14 that $I \succ J$ iff $J \succ K$ iff $K \succ I$; and $I \prec J$ iff $J \prec K$ iff $K \prec I$.

Thus, unless people are indifferent in all three choices, they must be intransitive in this design, according to this very general version of PRAM. By the same argument, PRAM must show recycling: if it is intransitive in this *IJK* design, it must show the opposite intransitive cycle in the *I′J′K′* design, as defined above.

PRAM implies RBI if the relative argument for S over R based on consequences, $\psi(y_S, y_R)$, has the property that it is independent of any branch that yields the same consequence for both alternatives, as it would if this relative argument followed an integration of contrasts as in Equations 1 or 10. From Expression 14, $S \succ R$ iff $\pi(b_R, b_S) > \psi(y_S, y_R)$. Note that in the RBI

designs, $\pi(b_S, b_R) = \pi(b_{S'}, b_{S'})$ because one event has the identical consequence, one event favors S, and one event favors R, and the same is true for the choice between S' and R'. So, the choice depends entirely on $\psi(y_S, y_R)$, which is the same as $\psi(y_{S'}, y_{R'})$, if the value of the common consequence does not interact with the contrasts of the consequences that differ. For example, if $\psi(y_S, y_R) = f[u(x_S), u(x_R)] + g[u(y_S), u(y_R)] + h[u(z_S), u(z_R)]$, where f, g, and h are skew symmetric functions, then $\psi(y_S, y_R) = \psi(y_{S'}, y_{R'})$. If $\pi(b_S, b_R) = \pi(b_{S'}, b_{S'})$ and $\psi(y_S, y_R) = \psi(y_{S'}, y_{R'})$, then $S \succ R$ iff $S' \succ R'$. Therefore, with these assumptions, PRAM falls in the same cell in Table 1 as Equations 1 and Equation 10, by predicting violation of transitivity and satisfaction of RBI.

The salience weighted utility model (Leland & Schneider, 2014) is similar to PRAM in that it includes contrasts in both probability and consequences. It differs from regret theory in that it does not satisfy the assumption of coalescing, so choices are expected to depend on the form of a choice problem. However, in the present experiments (in which choices are presented in canonical split form and the branches are equal in probability), this theory reduces to a special case of Expression 10 in which $\psi(x_i, y_i) = \mu(x_i, y_i)[u(x_i) - \underline{u}(y_i)]$, where $\mu(x_i, y_i)$ is the salience weighting function. For any $x > y > z$, it is assumed that $\mu(x, z) > \mu(x, y)$ and $\mu(x, z) > \mu(y, z)$; i.e., larger differences are more "salient" and receive greater weighting. With this assumption, this theory must be intransitive in the new *IJK* design and show the same pattern of intransitivity and recycling as regret theory. It also satisfies RBI, with or without the salience weighting.

Methods of Experiments 5 and 6

Experiments 5 and 6 included the new *IJK* experimental design for testing transitivity, mixed among other choice problems. Experiment 5 investigated whether people might adopt intransitive strategy in a longer experiment in which the same choice problems are presented repeatedly.

Perhaps with practice on the same choice problems, people might learn to adopt either advantage seeking or regret aversion as a way to decide between alternatives that are otherwise indifferent. Experiment 6 combined the main designs of Experiment 4 and 5 in a study with substantial monetary incentives. Perhaps such monetary incentives would intensify emotions of anticipated regret or advantage seeking that would lead to intransitive preferences. Alternately, such incentives might do the opposite.

The instructions, stimuli, and procedures of both experiments were similar to those of Experiment 4. The task was to choose between gambles displayed as in Figure 2.

Experiment 5

Experimental materials for one block of trials can be viewed at the following URL: http://ati-birnbaum.netfirms.com/Spr_2014/decision_diecidue_01.htm

There were 44 choice problems in each block of trials. When a block of trials was completed, the participant clicked a "submit" button to record the data and go on to the next block of trials. Each block of trials contained the same experimental choice problems, warmups and fillers in restricted random orders. Each participant was tested in the lab at a separate computer workstation, and worked at the task at her or his own pace for one hour.

Transitivity design: The transitivity design used I = (\$90, \$50, \$10), J = (\$10, \$90, \$50), K = (\$50, \$10, \$90). Each of the three choices was presented twice in each block, with positions counterbalanced: *IJ, JK, KI, JI, KJ, IK*.

RBI design: This design used choices between S = (\$5, x, y) and R = (\$5, \$10, \$90) and between S' = (\$95, x, y) and R' = (\$95, \$10, \$90). There were seven levels of (x, y) = (\$15, \$25), (\$20, \$30), (\$25, \$35), (\$30, \$40), (\$35, \$45), (\$40, \$50), and (\$45, \$55). These 2 X 7 = 14 choice problems were presented twice in each block, with order counterbalanced, making 28 choice

problems per block.

Dominance and filler designs: There were 10 additional trials, used as warmups and fillers. There were three choice problems testing transparent dominance, such as ($15, $40, $60) versus ($20, $45, $65). There were also two choice problems that pitted advantage seeking against regret aversion: *D* = ($70, $80, $10) versus *E* = ($20, $30, $90) and *D′* = ($35, $40, $10) versus *E′* = ($30, $35, $90). An extremely regret averse person would choose *E* and *E'*; an extreme advantage seeker would choose *D* and *D′,* but a person with tradeoffs might easily switch from *D* to *E′.* These two choice problems and the three tests of dominance were presented in counterbalanced positions, creating 10 trials per block.

There were 100 undergraduates tested in the lab; 53 were female and 95 were 22 years of age or younger. Of these, 12 participated on a second day for an additional hour. On average, participants completed 11.33 blocks of trials. Instructions specified that at the end of the semester, three participants would be randomly selected to play one of their chosen gambles for real money prizes. These were played out, as promised, and three people received prizes. The sample sizes in this study and in Experiment 6 were fixed in advance.

Experiment 6

In Experiment 6, every participant played out a gamble, so incentives were higher than in Experiment 5. Participants were recruited from a Website (advertised in flyers distributed at university libraries in Paris) to participate in studies for pay in the lab. Each participant was recruited to serve for one hour at a flat rate of 5 Euros; however, at the experiment, participants were told that immediately following the session, every participant would play one of her or his chosen gambles by drawing a colored chip from an urn to determine an additional prize as high as 95 Euros (19 times the standard hourly rate).

The instructions gave examples including cases of dominance in which a poor decision would be regretted, and the term "regret" was used to describe how a participant would feel if she or he made a decision that resulted in a smaller prize for a given color. Following the instructions, which included examples of how a gamble would be played out later, participants were reminded that a good prize would depend on both good decisions and good luck.

Participants were required to complete four blocks of trials, with two blocks for each of two experimental designs. Design 1 was a slightly revised version of Experiment 4 (42 trials/block), in which only warmup and filler trials were altered; Design 2 was a revised version of Experiment 5 (44 trials/block), including *IJK* choices of Experiment 5 plus three choice problems presented in counterbalanced orders (six trials): $I' = (15, 35, 85)$, $J' = (85, 15, 35)$, $K' = (35, 85, 15)$, presented in both counterbalanced orders. Six trials were deleted from Experiment 5: four trials from the RBI design for $(x, y) = (15, 25)$, and two trials testing dominance. All other choice problems remained the same as in Experiments 4 and 5, except prizes were to be paid in Euros rather than dollars (at the time, 1 Euro was worth 1.36 USD). The blocks were presented in the order: Design 1, Design 2, Design 1, Design 2.

The instructions (in French), as well as the experimental trials can be viewed at the following URLs:

http://ati-birnbaum.netfirms.com/e12_14/regret_d1_01.htm

http://ati-birnbaum.netfirms.com/e12_14/regret_d2_01.htm

The participants were 50 French-speaking people (66% female, 58% older than 22 years, and all were less than 32 years). All except 3 were French citizens, of whom all but two were born in France. These participants were more highly educated, on average, than lower division college undergrads tested in Experiments 1-3 and 5; 32 were college graduates, of whom 20 had post-

graduate education, including 13 with Master's degrees and 1 with a Ph.D.

There were about 4 participants per session. Each participant worked independently at a separate computer workstation. Following each session, all participants witnessed each person in the session playing out his or her chosen gamble.

The participant rolled multisided dice to determine which trial would be played; a printout of the participant's responses was consulted, and the participant drew a colored chip from an urn to determine the prize, according to the decision he or she had made on that trial. This procedure generated great interest and excitement, with celebrations and exultations of joy by those who won more than 50 Euros and expressions of disappointment by those who won 40 Euros or less (the expected value of all prizes (EV of a randomly chosen prize) was 46.87 Euros (variance = 111.28), but a participant could improve upon that by simply conforming to transparent dominance (which this sample did quite well), and one could improve beyond that by conforming to EV as well (which most did not). In this study, conforming to or violating transitivity would have no effect on EV of the take home prize, since the EVs of the gambles in the transitivity designs were fixed. The empirical mean prize was 58.1 Euros, on top of the 5 Euro payment. The mean prize is significantly higher than a randomly chosen prize ($z = 7.58$, $p < .01$).

Results of Experiments 5 and 6

Design 1 of Experiment 6 was intended to replicate the *ABC, A'B'C', and A''B''C''* designs of Experiment 4. Simplified results are presented in Table 12, which analyzes cases where people made the same response patterns on both replicates of the three choice problems testing transitivity within a block of trials. Each entry in the first 8 rows of data is the percentage of consistent response patterns of each type. Table 12 shows that in all three designs of Experiments 4 and 6, the most commonly repeated (consistent) response pattern was the transitive pattern *112*,

which comprised more than half of all cases where participants repeated the same pattern (53% to 72%).

Insert Table 12 about here.

The next to last row of Table 12, labeled "% repeats" shows the percentage of all response patterns that were consistent (where participant made the same responses on both replicates for all three choice items in a block). The last row of Table 12 shows the number of completed response patterns in each study.

For example, in the *ABC* design of Experiment 6, the number of response patterns was only 96 (instead of 100 = 50 participants X 2 blocks) because in four cases, a participant failed to respond to one of the 6 choice items in a block (3 choice problems X 2 repetitions). Of these 96 complete response patterns, participants repeated the same response pattern on both repetitions within a block on 49% of the patterns; and 2% of those 49% (1 instance) was the *111* pattern predicted by regret theory.

Detailed results for the *ABC, A'B'C', and A''B''C''* designs are presented in the Online supplementary materials (SM) for this article. Tables SM.7, SM.8, and SM.9 for Experiment 6 can be compared with Tables SM.3, SM.4, and SM.6 for Experiment 4, respectively. In all six tables (both experiments), the largest frequency occurs for the repeated transitive pattern, *112 112*. All three tables of Experiment 6 show relatively high frequency for patterns *221 221* and *212 212* as well as *112 212* and *212 112*, similar to results from Experiments 3 and 4.

Unlike Experiments 3 and 4, however, Table 12 (and Tables SM.7, SM.8, and SM.9) do not show any evidence of the infrequent but systematic intransitivity that appeared in Experiments 3 and 4 of the type predicted by advantage seeking or majority rule. Table 12 shows that in Experiment 4, about 5% of the consistent response patterns showed the intransitive pattern *222* in

the *ABC* design, and about 5% showed the intransitive pattern *111* in the *A′B′C′* design, consistent with advantage-seeking and majority rule; however, Table 12 shows that this evidence of infrequent but systematic intransitivity was not observed in Experiment 6. This small difference might represent something about the participants, incentives, or procedures in Experiment 6; however, the sample size in Experiment 6 was so small (n = 50) that it would not be too surprising to find almost no one showing evidence of intransitivity, even if both groups were random samples from the same populations. [If the probability of a person showing intransitivity is p = .06 and n = 50, the binomial probability to observe 1 or fewer cases out of 50 is 0.19.] Even if this small difference between studies is real, however, it is interesting that Experiment 6 showed even less evidence of intransitivity than Experiment 4.

Referenced Tables SM.7, SM.8, and SM.9 about here.

The last three columns of Table 12 show results from the *IJK* designs for Experiments 5 and 6 and the *I′J′K′* design of Experiment 6. Recall that I = (90, 50, 10), J = (10, 90, 50), and K = (50, 10, 90). In this design, Equation 13 with regret aversion (Expression 12a) must produce the intransitive pattern *111*, and Equation 13 with advantage seeking (Expression 12b) must produce the intransitive cycle, *222*. These predictions reverse in the permuted, *I′J′K′* design. Instead, examining the percentages of consistent response patterns, Table 12 shows that intransitive patterns comprise 8 to 23% of the consistent patterns. The highest incidence (18% for the *222* response pattern in Experiment 5) agrees with advantage seeking, as in Experiments 3 and 4; yet this same pattern is rare in Experiment 6. The highest incidence of intransitivity in Experiment 6 occurred for the *111* pattern in the *IJK* design, where 10% of the 42% of consistent patterns (4 instances out of 100) were observed, matching the predictions of regret aversion.

The most commonly repeated patterns in the *IJK* design of both Experiments 5 and 6, is the

transitive pattern, *122*. This pattern is consistent with the rule to choose that gamble that gives the higher prize if Red (the color listed on the left) is drawn from the urn. In the *I'J'K'* design the most frequently repeated pattern is *211*, which is also consistent with the same rule. Also prevalent are the patterns *221* in the *IJK* design and *112* in *the I'J'K'* design, which are consistent with the rule to choose the gamble with the higher consequence for drawing a Blue marble (rightmost position). These two transitive response patterns combined account for more than half of all repeated patterns in all three columns. Therefore, it appears that instead of adopting regret or advantage seeking as a secondary strategy to resolve choices in the new "desperate" design, the majority of participants adopted a transitive strategy.

Detailed results of the *IJK* and *I'J'K'* designs of Experiments 5 and 6 are presented in Tables SM.10, SM.11, and SM.12. These tables reveal unusual behavior in these new designs not seen in previous studies: Some participants simply chose the first gamble on these trials, which due to the counterbalancing procedure, results in the inconsistent pattern, *111 222*; this response pattern was either the most frequent response pattern or the second most frequent pattern in all three tables (upper right corner of Tables SM.10, SM.11, and SM.12).

Referenced Table SM.10, SM.11, and SM.12 about here.

The data for *IJK* design also show more "scatter", as if some people were choosing randomly or arbitrarily on these trials. In Experiment 5 (Table SM.10), repeated patterns (sum of entries on the diagonal) account for 364 cases, which sum to only 33% of all completed response patterns (1118 completed of possible 1133). Table 12 shows that repeated patterns accounted for 48%, 56%, 52%, 49%, 60%, and 60% in the *ABC, A'B'C' and A''B''C''* designs in Experiments 4 and 6, respectively. In Experiment 6, repeated patterns were only 42% and 37% in the *IJK* and *I'J'K'* designs.

In Experiment 6, one person was found who conformed to regret aversion on all 24 choice problems in the *IJK* and *I′J′K′* designs, who accounted for two of the four instances of the pattern *111 111* in Table SM.11 and two of the three cases in Table SM.12 of *222 222*. But that same person did not show any intransitive response patterns in either the *ABC* or *A′B′C′* designs (Tables SM.7 and SM.8), showing instead the *112* pattern in three of the four tests each in these designs.

Referenced Tables SM.11 and SM.12 about here.

Because the *IJK* experimental design requires intransitive behavior from any person following Equation 10 who shows either regret aversion or advantage seeking, and because there was very little else left for participants to do, the failure to observe intransitive preferences in this design must be regarded as a refutation of Equation 10. Because PRAM implies that people should either be intransitive in the *IJK* design or be indifferent, we can also reject PRAM.

Equation 10 implies that people should conform to RBI, apart from error. Because each item testing RBI was presented twice in each block of Experiments 5 and 6, we can use the *g*TET model to estimate the probabilities of errors and of true preferences. According to RBI, the true choice probabilities between $S = (5, x, y)$ and $R = (5, 10, 90)$ should be the same as the true choice probability between $S' = (95, x, y)$ and $R' = (95, 10, 90)$. Table 13 presents results for the four tests for which x is greater than or equal to 30. For example, the first two rows show that whereas 54% preferred $R = (5, 10, 90)$ over $S = (5, 30, 40)$ in the first block of trials in Experiment 5, 84% preferred $R' = (95, 10, 90)$ over $S' = (95, 30, 40)$ in the same block of trials. In all cases in Table 13, as the common consequence is increased, the choice proportion favoring the risky gamble increased, contrary to RBI but consistent with findings in previous studies.

Insert Table 13 about here.

Comparing the first block of trials to the last (Experiment 5), Table 13 shows that the

undergraduate sample became less risk averse in the *SR* choices in the last trial block and that the magnitude of the violations of RBI was reduced but still quite large. The error rates in Experiment 5 are clearly lower in the last block of trials than in the first. The error rates in Experiment 6, with the more highly educated and more highly motivated sample, are comparable to or even smaller than those in the last block of trials for the undergraduates in Experiment 5. The violations of RBI are quite substantial in Experiment 6, where this more highly educated sample was, on average, even more risk averse than the undergraduate sample of Experiment 5.

More details of the tests of RBI in Experiments 5 and 6 are presented in Tables SM.13, SM.14, and SM.15. All tests showed that people are more likely to choose the "risky" gamble when the common value is the highest than when it is the lowest value. For example, despite its higher expected value, only 40% and 29% preferred (5, 10, 90) over (5, 35, 45) in the first block of Experiment 5 and in Experiment 6, respectively. However, when the common prize on Red was changed from 5 to 95, the figures were 69% and 46% preferring the riskier and higher EV alternative (95, 10, 90) over (95, 35, 40). The sample with higher financial motivation showed even greater risk aversion (discrepancy from EV) than did the undergraduate sample with smaller financial motivation. All 24 tests of response independence in Tables SM.13-SM.15 were statistically significant, and all 24 $\chi^2(1)$ testing response independence were greater than the corresponding $\chi^2(1)$ testing *g*TET on the same data. In the last block of Experiment 5 and in Experiment 6, all $\chi^2(1)$ values testing error independence (testing *g*TET) were less than 2.0 and all $\chi^2(1)$ testing response independence were greater than 30.

Referenced Table SM.13, SM.14, and SM.15.

Tables 13 and SM.13-SM.15 add to the systematic violations of RBI observed in

Experiments 1-4. However, they go beyond previous results because they show that even with financial incentives, including considerable ones in Experiment 6, people systematically violate RBI. As noted in Table 1, violation of RBI not only refutes the most general form of Equation 10, it also disproves expected utility theory, constant-weight averaging model, and original prospect theory (with or without the editing rule of cancellation).

The fact that violations of RBI persist in Experiments 5 and 6 with monetary incentives should be of interest to economists as well as psychologists. The analysis shows that the violations cannot be attributed to random errors as defined in the *g*TET model. Finally, these violations occur with dependent gambles with a clear state space.

In sum, Experiments 5 and 6 show that even in a design where Equation 10, PRAM, and salience weighted utility must yield intransitive preferences, we do not observe many instances of intransitive preferences. A few people in Experiment 5 appear to show evidence of advantage-seeking, as in Experiments 3 and 4. Overall, however, the majority of participants in Experiments 5 and 6 were either transitive in these designs (e.g., choosing what is best under Red), responded arbitrarily (e.g., chose the first alternative), or chose randomly on these trials. Furthermore, both experiments show that people systematically violate RBI, even with monetary incentives, contrary to Equation 10, to salience weighted utility, and to the interpretation of PRAM in which relative arguments based on consequences are formed by comparing consequences for the same events.

General Discussion

These six experiments cannot be reconciled with the family of models in Table 1 that imply violations of transitivity and satisfaction of RBI. This part of Table 1 includes Equations 1 and 10 which allow for regret aversion, advantage seeking (including most probable winner theory and majority rule) the PRAM model (Loomes, 2010), and salience weighted utility (Leland &

Schneider, 2014).

The Search for Intransitive Preferences

Our search for violations of transitivity led to more and more restricted domains where more and more general models had less and less wiggle room to allow them to handle transitive as well as intransitive patterns.

Majority rule is a theory that allows no wiggle room; it must predict intransitive choices in all six experiments. Experiment 1 showed that most participants simply did not agree with its predictions, so the model can be rejected as a descriptive theory of how people choose between cities based on friend's advice. This experiment disproves majority rule when average value varies and led us to restrict our experimental designs in subsequent studies to choices with constant means.

Experiment 2 tested the proposition that if the mean ratings of the cities were kept constant, majority rule might be descriptive as a tiebreaker rule in this restricted domain. It was found that a small percentage satisfied the predictions but that most people did not agree with its predictions even in this restricted situation. So, majority rule can be rejected as a general descriptive model even when the mean ratings are fixed. This result contradicts the conclusions reached by Zhang, et al. (2006), who suggested that majority rule might be descriptive of how people make choices.

Experiments 3 and 4 tested a fairly general model (Equation 13) for choices between gambles that includes regret theory and advantage seeking (a generalization of majority rule) in a still more restricted situation in which expected values were kept the same in each choice problem and probabilities of the three events were known to be equal. Our experiments were devised to search for violations where regret theory (as fit to previous empirical data) predicted that we should find intransitive preference cycles and evidence of recycling. Surprisingly, we found that no one in those experiments conformed to those predictions; instead, most people were transitive, except for a

very small percentage of people who showed the opposite pattern of intransitivity and recycling from that predicted by regret theory.

Whereas our Experiments 3-4 were based on parameters of regret theory estimated in previous research (Bleichrodt, et al., 2010), Baillon, Bleichrodt, & Cillo (2014) recently used an even more elegant method: they fit regret theory to each individual and specifically tailored tests of transitivity for each person. Baillon, et al. (2014) found no evidence of intransitivity, despite these tailored designs that should have found it if regret theory were descriptive. Apparently, one cannot use the parameters of regret theory to predict from one set of choice problems to another, either from one study to the next or from one set of choices by a person to another set of choices by that same person.

Finally, in Experiments 5 and 6, we devised and tested a design where quite general expressions (Equations 1, 10, 13, and 14) must predict intransitive cycles. Even in this very constrained region in the space of possible choices, we found that most participants either continued to show transitive response patterns or responded arbitrarily by, for example, always choosing the first alternative in such choices. Because these models must imply intransitive cycles for all parameters (except when they reduce to EU), we must reject these models even for this highly restricted but friendly environment for intransitive models.

It is true that we demand a higher standard of evidence for transitivity than has been applied in the past. This higher standard requires showing that observed violations cannot be attributed to mixtures of transitive patterns or mixtures combined with random error, such as would occur if the same person sometimes makes different responses to repetitions of the same choice problem in the same block of trials. The higher standard for evidence also sets a higher standard for experimental design: the experiment should include a measurement of error for each choice problem. This does

not strike us as an unreasonable demand for a properly designed experiment. But these higher standards also mean that evidence of asymmetry of different types of violations does not count as meaningful evidence of intransitivity, since asymmetry is compatible with purely transitive preferences in the presence of error. Nor do tests of WST or the triangle inequality count as unambiguous tests of transitivity, since they could lead to wrong conclusions that can be avoided by proper analysis of proper experiments.

We do not think that the standard of evidence demanded is excessive. As noted in the results to Experiments 1 and 2, the TE model can in principle detect small percentages of people who show different types of transitive and intransitive patterns. Indeed, we detected small percentages who are consistent with intransitive models (Experiments 2, 3, and 4), but those percentages are so small that these theories are not useful empirical models since the vast majority of participants (more than 80% in each study) appear to be transitive. Because earlier methods for analyzing transitivity pool over response sequences and have no method to estimate error, they are not able to detect intransitivity in such cases. As examples in Birnbaum (2013) illustrate, they can even fail to find evidence of intransitivity when everyone was perfectly intransitive.

In Experiment 2, 3, and 4 small percentages were found who showed the type of intransitivity and recycling predicted by Equation 10 with advantage seeking. Therefore, it is not the standard of evidence or our method of analysis that should be blamed, but rather the behavior of the participants that has led to refutation of Expressions 1, 10 and 14 as useful descriptive models for the majority of participants.

We identified and tested a property that distinguishes intransitive models that we call recycling, which is implied by Expression 10, by PRAM, and by salience weighted utility. Recycling is the property that one can find stimuli that will show intransitive cycles that can be

reversed to produce opposite intransitive cycles by an appropriate permutation of the components. We found that, indeed, recycling was observed in that small percentage of people in Experiments 2-4 who were identified as intransitive. This small percentage showed the opposite pattern of behavior, however, from that predicted by regret aversion (Expression 12a); this percentage showed evidence of advantage seeking (Expression 12b).

It is perhaps ironic that out of 1224 participants in these studies, we found only one participant (in Experiment 6) who showed the pattern of intransitive behavior predicted by regret aversion, and that one person showed this behavior only in the narrow, *IJK* and *I'J'K'* designs (where all three consequences were the same in both alternatives of each choice). Furthermore, that same individual did not show intransitive behavior in the *ABC* and *A'B'C'* designs (where the means were the same).

Other recent studies of transitivity have also failed to reject this property (Birnbaum, 2010; Birnbaum & Bahra, 2012b; Birnbaum & Gutierrez, 2007; Birnbaum & Schmidt, 2008; Cavagnaro & Davis-Stober, 2014; Regenwetter, Dana, & Davis-Stober, 2010, 2011). Studies by Birnbaum and Gutierrez were designed to test intransitivity predicted by a lexicographic semi-order, as were other tests in Birnbaum (2010). Two studies by Birnbaum and Schmidt (2008) were devised to test regret theory and the most probable winner models, a generalization of majority rule. These recent studies have found that very few people repeat the same intransitive pattern on two replications of the same test. In other words, most violations that have been observed can be attributed to error rather than to true intransitivity.

For those participants who conformed to transitivity, we can retain the models in the lower left corner of Table 1, namely the configural weight averaging model and models like it (Luce & Marley, 2005; Marley & Luce, 2005). The finding that data are consistent with transitivity does not prove that transitivity is always satisfied, so such findings must be interpreted cautiously with

respect to general conclusions regarding transitivity: there might be some other cases not yet tested where intransitive preferences might be found. But these data do suggest that one would do best to set aside models based on Equations 1 and 10, and those seeking to test transitivity should pursue predictions of intransitivity by other models that have not yet been disproved.

Violations of Restricted Branch Independence

Systematic violations of RBI of the same type were observed in all six experiments. Violations of RBI refute Equations 1 and Expression 10, including cases of advantage seeking and regret aversion.

In reference to Table 1, violations of RBI require us to reject the constant weight averaging model (Anderson, 1974) including EU theory, "stripped" prospect theory (Edwards, 1954; Starmer & Sugden, 1993), and the cancellation rule of original prospect theory (Kahneman & Tversky, 1979). Recall that EU can be derived as the transitive special case of Expression 10. Violations of RBI rule out even this transitive special case of Equation 10.

Violations of RBI also violate salience weighted utility and PRAM (Equation 14), if it is assumed that the relative arguments based on consequences are represented by contrasts between consequences for corresponding events, as in Expression 10. See Loomes (2010, Footnote 19).

Violations of RBI remain consistent with configural weight models (Equation 5), including TAX and CPT, among others.

The "special" TAX model is a special case of Equation 5 for equally likely events; but it is also a more general model (than Equation 5) in that it describes choices between either independent or dependent gambles with unequal probabilities in which branches may be either coalesced or split (Birnbaum, 2008a). It is transitive and it predicts the type of violations of RBI that we observed in all six experiments. With its "prior" parameters, special TAX correctly predicted the most frequent

transitive response patterns observed in Experiments 1-4 in tests of transitivity. Equation 5 could also account for the most frequent transitive patterns observed in Experiments 5-6 if it allows the weights for the color-positions (*a*, *b*, and *c* for Red, White, and Blue) to differ.

Systematic violations of RBI have also been reported in choices between independent gambles by Birnbaum and Navarrete (1998), Birnbaum (1999, 2008a), and between dependent gambles by Birnbaum (2006) and Birnbaum and Bahra (2012a). Birnbaum and Beeghley (1997) found similar, systematic violations in judgments of value, as did Birnbaum and Zimmermann (1998). Because these violations are large and significant, and systematically show the same type of pattern as has been reported for other related tasks, it seems we can reject models that imply RBI as being generally descriptive for either choice or judgment. The fact that they occur in judgment, where gambles are presented one at a time in random order, as well as in choice suggests that they arise in the evaluation of gambles (*I* in Figure 3), rather than in comparisons (*C* in Figure 3) between gambles.

Although Equation 5 (including TAX) is the model that best describes the data of the majority of participants, it cannot describe results for those people who showed advantage-seeking intransitive cycles and recycles in Experiments 2-4. It seems plausible that these people use a transitive model such as TAX when choice problems allow it, and when the alternatives were indifferent to them (as when the means are held constant), they used majority rule or advantage seeking as a secondary, tiebreaking strategy, to decide between otherwise indistinguishable alternatives. But apparently, most people (including those in Experiments 5 and 6) found other ways to resolve such otherwise difficult choice problems without violating transitivity.

CPT is also a special case of Equation 5, but the pattern of violations of RBI found in this study is of the opposite type from what is predicted by the inverse-S weighting function of CPT

(Tversky & Kahneman, 1992), which is required in that theory to account for well-established phenomena, such as the Allais paradoxes and risk-seeking for small probabilities.

This pattern of violations of RBI (that contradicts CPT with inverse-S) had already been observed in 36 previously published studies of RBI reviewed by Birnbaum (2008a, Tables 14, 18, and 19). Birnbaum and Bahra (2012a) collected many repetitions from each participant and found 74 out of 102 participants had violations of RBI that were statistically significant when tested within- person. Only one of these 74 had violations that were consistent with predictions of CPT with an inverse-S weighting function, and that one participant could not be represented by CPT because that person violated first order stochastic dominance in 30 of 32 tests. Therefore, the systematic violations of RBI observed here add to a substantial literature contradicting the inverse-S weighting function of CPT. Because the inverse-S function is required in that theory to account for other phenomena, it is reasonable to conclude that CPT is simply false, a conclusion confirmed by tests of other critical properties of that model (Birnbaum, 2008a).

Violations of critical properties of CPT cannot be explained by changing the weighting function or value function, but lead to rejection of all forms of Rank and Sign-Dependent utility (Luce, 2000; Luce & Fishburn, 1991, 1995), including CPT. These critical properties include first order stochastic dominance, upper and lower cumulative independence, upper tail independence, and gain-loss separability (Birnbaum, 2008a, 2008b; Birnbaum & Bahra, 2007; Wu, 1994; Wu & Markle, 2008). These critical violations, also called "new paradoxes," imply that CPT can no longer be considered a viable descriptive theory of risky decision-making (Birnbaum, 2008a). In summary, although CPT implies transitive preferences and can violate RBI, two findings of the present studies, it predicts the wrong pattern of violation of RBI and it has been rejected as a descriptive model in a number of empirical studies of its critical properties.

Revising PRAM and Salience Weighted Utility

As noted in Loomes (2010), one might view the approach of PRAM as a framework that includes all models of decision making, including transitive and intransitive models, rather than as a particular instance of such a model. The particular instance favored by Loomes (2010) is one in which probabilities and consequences are contrasted prior to integration, as in Figure 4 that lead to intransitive preferences, rather than the transitive models such as in Figure 3.

How could one revise this form of PRAM to account for our results? If consequences for the same events are compared between gambles (Loomes, 2010, Footnote 19), then common event-consequence branches will be cancelled. Thus, in tests of RBI, the choice between S and R must be the same as the choice between S' and R'. If consequences are ranked first, however, and weighted by rank as in the configural weight models of Equation 5, and then contrasted, then RBI can be violated and transitivity satisfied in Experiments 3 and 4. This modification introduces configural interactions within gambles (ranking and weighting of branches), before the branch values are contrasted. If consequences are ranked before being compared (as in the "control" designs of Exp. 4 and 6), this modified PRAM model would not violate transitivity in Experiments 3 and 4 (but they would be indifferent in the desperate designs of Experiments 5 and 6). Thus, this modified version of PRAM, allowing configural effects within gambles prior to comparisons between gambles, remains compatible with two major findings reported here: transitivity satisfied and RBI violated. But even this modified version of PRAM is refuted by Birnbaum's (2008a) "new paradoxes", however, since even this revision would not account for violations of coalescing and properties that follow from coalescing, including the dissection of Allais paradoxes (Birnbaum, 2004a).

Salience weighted utility (Leland & Schneider, 2014) violates coalescing and might therefore be better able to take advantage of the suggestion that branches be ranked by their

consequences rather than compared by events. However, it is unclear how to revise this model to allow for violations of RBI and also account for "new paradoxes" such as violations of upper tail independence and violations of branch splitting independence, cases where there are unequal numbers of branches in the gambles being compared (Birnbaum, 2007, 2008a).

TE Models

Tests of the TE models and of response independence in this study are consistent with previous findings: the TE model fits the data better than models that assume response independence. According to the TE model, violations of independence arise from mixtures of different preference patterns. In the TE analysis of group data (*g*TET), mixtures arise because of individual differences among participants. In the analysis of individual data (*i*TET), a mixture arises because the individual systematically changes preferences over time. In either *i*TET or *g*TET, response independence will follow when there is only one true response pattern in the mixture. This assumption might be a good approximation for the choices in this study testing "transparent" dominance, but not in any case of our choice problems testing transitivity. In most cases, the TE model fit the data well, though a few cases of significant deviations were found.

Concluding Comments

Much work has been accomplished on the theoretical implications and analysis of majority rule and regret theory, but the failures of theory to correctly predict new, potentially diagnostic results raise questions about the descriptive adequacy of this family of theories. It is worth repeating that previous results cited as evidence for intransitive cycles consistent with regret aversion and PRAM were not really diagnostic since they were based on methods that could lead to wrong conclusions. A TE model in which no one is truly intransitive could easily show the asymmetry of different types of intransitive cycles, previously cited as evidence of regret cycles

(Birnbaum & Schmidt, 2008).

In sum, our data indicate that most of the people we tested were consistent with transitive models, but a small minority showed systematic intransitivity and recycling. However, most of those people who did appear to recycle seem to have no regrets.

Appendix A. True and Error Model for Transitivity and Recycling

Suppose that different people may have different true preferences, but the same person might make different responses to the same choice problem within in a block of trials due to random errors. In our tests of transitivity, each choice was presented twice in each block, which allows us to estimate error rates from preference reversals between two presentations of the same choice problem within blocks. In this study, positions of the alternatives were counterbalanced, so the participant had to push opposite buttons on different trials to be consistent.

Let e_1 represent the error rate for choice problem, *A versus B*. The probability of choosing *A* over *B* in one presentation and *B* over *A* in the other presentation is given as follows:

$$P(AB) = p(e_1)(1 - e_1) + (1 - p)(e_1)(1 - e_1) = (e_1)(1 - e_1)$$

Where $P(AB)$ is the predicted probability of choosing *A* over *B* on one trial and on a separate trial, separated by intervening trials, of choosing *B* over *A* when the same alternatives were presented in counterbalanced order; where p is the true probability of preferring *A* over *B*. The opposite reversal of preference has the same probability, so $P(AB) + P(BA) = 2(e_1)(1 - e_1)$. For example, if there are 32% reversals of preference between repetitions of the *AB* choice, the error rate is $e_1 = 0.2$.

In Experiments 3 and 4, we theorize up to 10 possible true response patterns regarding the choices among *A*, *B*, *C* and *A'*, *B'*, and *C'*. There are six transitive patterns, which are denoted *112*, *121*, *122*, *211*, *212*, and *221*, where *112* represents the transitive response pattern of choosing *A* over *B*, *B* over *C*, and *A* over *C* in the *AB*, *BC*, and *CA* choices, and where the corresponding

preferences hold for *A′*, *B′*, and *C′*. We assumed for simplicity that if a person is transitive, that person would also have the same true preferences for the corresponding choices in the *A′B′C′* design as in the *ABC* design; for example, for the TAX model with its prior parameters, aside from error, the predicted response pattern is *112 112* for the six choices. The probabilities of the six transitive response patterns are denoted: p_{112}, p_{121}, p_{122}, p_{211}, p_{212}, and p_{221}.

In addition to the six transitive patterns, the model allowed four intransitive and recycling response patterns: According to regret theory as fit to previous data (R), a person should have the intransitive response patterns of *111* in the *ABC* design and *222* in the *A′B′C′* design, denoted *111 222*, showing both intransitivity and recycling. Let p_{R} represent the probability that a person is governed by these preference patterns of regret theory (regret aversion).

A person who satisfies majority rule (MR), however, should have the opposite patterns of intransitivity and recycling; i.e., *222* and *111*, respectively. Let p_{MR} represent the probability that a person has these preference patterns. It is also possible that a person might have the same intransitive pattern in both designs (*111 111* or *222 222*, respectively) with probabilities of p_{111} and p_{222}, respectively.

In Experiment 4, in the *A″B″*, *B″C″*, and *C″A″* choices, neither MR nor R theory implies recycling. In that case, we assumed that the sum of the other eight response patterns would be 1 ($p_{112} + p_{121} + p_{122} + p_{211} + p_{212} + p_{221} + p_{111} + p_{222} = 1$).

The probability of an error is allowed to differ for different choice problems. Initially, we allowed that choices in the different permutation sub-designs might have different rates of errors: e_1, e_2, e_3, e'_1, e'_2, and e'_3, respectively in the choices *AB*, *BC*, *CA*, *A′B′*, *B′C′*, and *C′A′*, respectively. In Experiment 4, errors in the *A″B″*, *B″C″*, and *C″A″* choices of the control design are denoted, f_1,

f_2, and f_3, respectively. (Based on analysis of the data, we later concluded that the error rates can be assumed to be independent of permutations.)

The probability to show the (transitive) response pattern, *112*, in the *ABC* design is given as follows:

$$P(112) = (p_{111} + p_R)(1 - e_1)(1 - e_2)(e_3) + p_{112}(1 - e_1)(1 - e_2)(1 - e_3) +$$
$$+ p_{121}(1 - e_1)(e_2)(e_3) + p_{122}(1 - e_1)(e_2)(1 - e_3) +$$
$$+ p_{211}(e_1)(1 - e_2)(e_3) + p_{212}(e_1)(1 - e_2)(1 - e_3) +$$
$$+ p_{221}(e_1)(e_2)(e_3) + (p_{222} + p_{MR})(e_1)(e_2)(1 - e_3)$$

where $P(112)$ is the probability to show this response pattern. There are seven other equations for the other seven possible response patterns. The probability to repeat this response pattern on both repetitions of a block is given by the same equation, except that each of the error terms, e or $1 - e$, is squared.

Similarly, one can write equations for the eight possible response patterns in the *A'B'C'* design, including the probabilities of showing each possible response pattern on two repetitions of a design. In this design, MR implies the pattern *111* and R implies the pattern *222*, and the error terms are e'_1, e'_2, and e'_3. In the *A''B''C''* design of Experiment 4, the equations are similar, but neither MR nor R imply intransitivity, and the error terms are f_1, f_2, and f_3.

The probability to show the intransitive and recycling response pattern predicted by MR model, *222* and *111* in both *ABC* and *A'B'C'* designs, respectively, denoted $P(222\ 111)$, is as follows:

$$P(222\ 111) = p_R(e_1)(e_2)(e_3)(e'_1)(e'_2)(e'_3)$$
$$+ p_{111}(e_1)(e_2)(e_3)(1 - e'_1)(1 - e'_2)(1 - e'_3)$$

$$+ p_{112}(e_1)(e_2)(1-e_3)(1-e'_1)(1-e'_2)(e'_3)$$

$$+ p_{121}(e_1)(1-e_2)(e_3)(1-e'_1)(e'_2)(1-e'_3)$$

$$+ p_{122}(e_1)(1-e_2)(1-e_3)(1-e'_1)(e'_2)(e'_3)$$

$$+ p_{211}(1-e_1)(e_2)(e_3)(e'_1)(1-e'_2)(1-e'_3)$$

$$+ p_{212}(1-e_1)(e_2)(1-e_3)(e'_1)(1-e'_2)(e'_3)$$

$$+ p_{221}(1-e_1)(1-e_2)(e_3)\ (1-e'_1)(1-e'_2)(1-e'_3)$$

$$+ p_{222}(1-e_1)(1-e_2)(1-e_3)(e'_1)(e'_2)(e'_3)$$

$$+ p_{\mathrm{MR}}(1-e_1)(1-e_2)(1-e_3)(1-e'_1)(1-e'_2)(1-e'_3)$$

There are a 63 other equations like the above, each of which is the sum of ten terms (representing the ten theoretical true preference patterns), for the 64 possible relative frequencies of response patterns in *ABC* X *A'B'C'* cross-tabulation (e.g., Table 9).

In fitting the model, we constrained the model to approximate the 64 possible response patterns in each of three, 8 X 8 cross-tabulation tables: *ABC* X *ABC*, *A'B'C'* X *A'B'C'*, and *ABC* X *A'B'C'.* In Experiment 4, the model was constrained to fit these three tables as well as the *A''B''C''* X *A''B''C''.* An index of fit was calculated in each of these tables, between the predicted frequencies and observed frequencies of the different response patterns, and the three indices for Experiment 3(or four, in Experiment 4) indices were summed to produce an overall index of fit.

For the 64 cells in each 8 X 8 cross-tabulation table, we compute the G^2 statistic as follows:

$$G^2 = -2\sum\sum F_{ij}\,[\ln(F_{ij}/P_{ij})],$$

Where F_{ij} and P_{ij} are the observed and predicted frequencies of each (of 64) response patterns, given the TE model. This statistic is theoretically distributed as a Chi-Square, and its value is

usually close to the value of a conventional Chi-Square calculation. It is regarded as a better index when cell frequencies are small.

When fitting Experiment 3, we sought parameters to minimize sum of the three G^2 indices, defined on Tables 7, 8, and 9), which constrains the model to approximate all three, 8 X 8 arrays of data. These three indices contain redundant information, so one cannot consider these sums to follow a known sampling distribution. In Experiment 4, the overall index is the sum of the four indices, defined on Tables SM.3, SM.4, SM.5, and SM.6.

The general model of Experiment 3 has 10 parameters representing relative frequencies of the 10 response patterns (which are assumed to sum to 1) and six error terms for the six choice problems in the *ABC* and *A'B'C'* designs, respectively. The purely transitive model is a special case of the general TE model that assumes that the four components that allow intransitive response patterns are all fixed to zero: $p_{111} = p_{222} = p_R = p_{MR} = 0$, and the other six transitive patterns are free and sum to 1 (using five degrees of freedom).

In fitting the models, it was found that the error terms could be constrained as follows: $e_1 = e'_1 = f_1$, $e_2 = e'_2 = f_2$, and $e_3 = e'_3 = f_3$, with only a tiny loss of fit. In none of our analyses did the difference in fit between these constrained models amount to more than a trivial improvement. In addition, three of the intransitive terms could be set to zero without noticeable loss of fit, $p_{111} = p_{222} = p_R = 0$. However, constraining $p_{MR} = 0$ produced a relatively larger drop in fit. This more substantial loss of fit (Table 10), combined with the systematic effects observable in the tables and the individuals whose responses perfectly matched the MR model predictions led us to conclude that a small proportion of participants can be represented by MR. In contrast, no one fit the regret model perfectly, the model was not noticeably improved by estimating $p_R > 0$, and best-fit estimates

of $p_R \leq 0.01$

The correlations between the predicted and observed frequencies in Tables 7, 8, and 9 of Experiment 3 for the MR model with constrained errors were .951, .967, and .942, respectively. The correlations between predicted and obtained frequencies for the same model in Experiment 4 were .989, .985, .982, and .986 for Tables SM.3, SM.4, SM.5, and SM.6, respectively.

These TE models are more accurate than any model that assumes responses are independent. The corresponding correlations for the theory that responses are independent (that frequencies of response patterns can be reproduced from products of the marginal binary choice proportions) are .620, .780, and .708 for Tables 7, 8, and 9 in Experiment 3 and .793, .896, .860, and .848 for Tables SM.3, SM.4, SM.5, and SM.6 in Experiment 4, respectively). People are much more consistent in their choices than predicted by the assumption of response independence. For example, response independence predicts that people should have agreed with their own patterns only 16% and 20% of the time in Tables 7 and SM.3 (sum of the diagonal entries) but the actual figures were 42% and 48%, respectively. The TE (MR model with constrained errors) predicted that a person would show the same pattern of responses on both replicates within a block 37% and 48%, respectively. For more on TE models compared with the theory of independent responses, see Birnbaum (2013).

References

Anderson, N. H. (1974). Information integration theory: A brief survey. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (pp. 236-305). San Francisco: W. H. Freeman.

Baillon, A., Bleichrodt, H., & Cillo, A. (2014). A tailor-made test of intransitive choice. *Manuscript*.

Bernoulli, D. (b. L. S. (1954). Exposition of a new theory on the measurement of risk.

Econometrica, 22, 23–36. (Translation of Bernoulli, D., 1738.)

Birnbaum, M. H. (1974). The nonadditivity of personality impressions. *Journal of Experimental Psychology, 102*, 543-561.

Birnbaum, M. H. (1982). Controversies in psychological measurement. In B. Wegener (Eds.), *Social attitudes and psychophysical measurement* (pp. 401-485). Hillsdale, N. J.: Erlbaum.

Birnbaum, M. H. (1999a). Paradoxes of Allais, stochastic dominance, and decision weights. In J. Shanteau, B. A. Mellers, & D. A. Schum (Eds.), *Decision science and technology: Reflections on the contributions of Ward Edwards* (pp. 27-52). Norwell, MA: Kluwer Academic Publishers.

Birnbaum, M. H. (1999b). Testing critical properties of decision making on the Internet. *Psychological Science, 10*, 399-407.

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

Birnbaum, M. H. (2004b). Tests of rank-dependent utility and cumulative prospect theory in gambles represented by natural frequencies: Effects of format, event framing, and branch splitting. *Organizational Behavior and Human Decision Processes, 95*, 40-65.

Birnbaum, M. H. (2005). A comparison of five models that predict violations of first-order stochastic dominance in risky decision making. *Journal of Risk and Uncertainty, 31,* 263-287.

Birnbaum, M. H. (2006). Evidence against prospect theories in gambles with positive, negative, and mixed consequences. *Journal of Economic Psychology, 27,* 737-761.

Birnbaum, M. H. (2007). Tests of branch splitting and branch-splitting independence in Allais paradoxes with positive and mixed consequences. *Organizational Behavior and Human*

Decision Processes, 102, 154-173.

Birnbaum, M. H. (2008a). New paradoxes of risky decision making. *Psychological Review, 115*, 463-501.

Birnbaum, M. H. (2008b). New tests of cumulative prospect theory and the priority heuristic: Probability-outcome tradeoff with branch splitting. *Judgment and Decision Making, 3,* 304-316.

Birnbaum, M. H. (2010). Testing lexicographic semi-orders as models of decision making: Priority dominance, integration, interaction, and transitivity. *Journal of Mathematical Psychology, 54*, 363-386.

Birnbaum, M. H. (2011). Testing mixture models of transitive preference: Comment on Regenwetter, Dana, and Davis-Stober (2011). *Psychological Review, 118*, 675-683.

Birnbaum, M. H. (2012a). A statistical test of the assumption that repeated choices are independently and identically distributed. *Judgment and Decision Making, 7*, 97-109.

Birnbaum, M. H. (2012b). True and error models of response variation in judgment and decision tasks. *Workshop on Noise and Imprecision in Individual and Interactive Decision-Making*, University of Warwick, U.K., April, 2012.

Birnbaum, M. H. (2013). True-and-error models violate independence and yet they are testable. *Judgment and Decision Making, 8,* 717-737.

Birnbaum, M. H., & Bahra, J. P. (2007). Gain-loss separability and coalescing in risky decision making. *Management Science, 53*, 1016-1028.

Birnbaum, M. H., & Bahra, J. P. (2012a). Separating response variability from structural inconsistency to test models of risky decision making. *Judgment and Decision Making, 7*, 402-426.

Birnbaum, M. H., & Bahra, J. P. (2012b). Testing transitivity of preferences in individuals using linked designs. *Judgment and Decision Making, 7,* 524-567.

Birnbaum, M. H., & Gutierrez, R. J. (2007). Testing for intransitivity of preferences predicted by a lexicographic semiorder. *Organizational Behavior and Human Decision Processes, 104*, 97-112.

Birnbaum, M. H., & LaCroix, A. R. (2008). Dimension integration: Testing models without trade-offs. *Organizational Behavior and Human Decision Processes, 105*, 122-133.

Birnbaum, M. H., & McIntosh, W. R. (1996). Violations of branch independence in choices between gambles. *Organizational Behavior and Human Decision Processes, 67,* 91-110.

Birnbaum, M. H., & Navarrete, J. B. (1998). Testing descriptive utility theories: Violations of stochastic dominance and cumulative independence. *Journal of Risk and Uncertainty, 17*, 49-78.

Birnbaum, M. H., & Schmidt, U. (2008). An experimental investigation of violations of transitivity in choice under uncertainty. *Journal of Risk and Uncertainty, 37,* 77-91.

Birnbaum, M. H., & Stegner, S. E. (1979). Source credibility in social judgment: Bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology, 37*, 48-74.

Birnbaum, M. H., & Sutton, S. E. (1992). Scale convergence and utility measurement. *Organizational Behavior and Human Decision Processes, 52*, 183-215.

Birnbaum, M. H., & Zimmermann, J. M. (1998). Buying and selling prices of investments: Configural weight model of interactions predicts violations of joint independence. *Organizational Behavior and Human Decision Processes, 74*(2), 145-187.

Birnbaum, M. H., Patton, J. N., & Lott, M. K. (1999). Evidence against rank-dependent utility theories: Tests of cumulative independence, interval independence, stochastic dominance,

and transitivity. *Organizational Behavior and Human Decision Processes, 77,* 44-83.

Blavatskyy, P. R. (2006). Axiomatization of a preference for most probable winner. *Theory and Decision, 60*, 17-33.

Bleichrodt, H., Cillo, A., & Diecidue, E. (2010). A quantitative measurement of regret theory. *Management Science, 56*, 161-175.

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Choices without tradeoffs. *Psychological Review, 113*, 409-432.

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.

Cha, Y., Choi, M., Guo, Y., Regenwetter, M., & Zwilling, C. (2013). Reply: Birnbaum's (2012) statistical tests of independence have unknown Type-I error rates and do not replicate within participant. *Judgment and Decision Making, 8,* 55–73.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin, 51*, 380-417.

Fishburn, P. C. (1982). Non-transitive measureable utility. *Journal of Mathematical Psychology, 26*, 31-67.

González-Vallejo, C. (2002). Making trade-offs: A probabilistic and context-sensitive model of choice behavior. *Psychological Review, 109*, 137-155.

Johnson, J. G., & Busemeyer, J. R. (2005). A dynamic, computational model of preference reversal phenomena. *Psychological Review, 112*, 841-861.

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

Leland, J. W. (1998). Similarity judgments in choice under uncertainty: A re-interpretation of the predictions of regret theory. *Management Science, 44*, 659-672.

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

Leland, J.W., & Schneider, M. (2014). Salience weighted utility and the framing of decisions. *Manuscript*.

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology, 89,* 46–55.

Loomes, G. (1988). When actions speak louder than prospects. *American Economic Review, 78*, 463-470.

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

Loomes, G. & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal, 92*, 805-824.

Loomes, G., Starmer, C., & Sugden, R. (1991). Observing violations of transitivity by experimental methods. *Econometrica, 59*, 425-440.

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

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

Luce, R. D., & Fishburn, P. C. (1991). Rank- and sign-dependent linear utility models for finite first-order gambles. *Journal of Risk and Uncertainty, 4*, 29–59.

Luce, R. D., & Fishburn, P. C. (1995). A note on deriving rank-dependent utility using additive joint receipts. *Journal of Risk and Uncertainty, 11*, 5–16.

Luce, R. D., & Marley, A. A. J. (2005). Ranked additive utility representations of gambles: Old and new axiomatizations. *Journal of Risk and Uncertainty, 30*, 21–62.

Marley, A. A. J., & Luce, R. D. (2005). Independence properties vis-à-vis several utility representations. *Theory and Decision, 58*, 77-143.

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

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

Russo, E., & Dosher, B. A. (1983). Strategies for multiattribute binary choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 676-696.

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

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

Starmer, C., & Sugden, R. (1993). Testing for juxtaposition and event-splitting effects. *Journal of Risk and Uncertainty, 6*, 235-254.

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

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

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5,* 297-323.

Wu, G. (1994). An empirical test of ordinal independence. *Journal of Risk and Uncertainty, 9*, 39–60.

Wu, G., & Markle, A. B. (2008). An empirical test of gain-loss separability in prospect theory. *Management Science, 54*, 1322-1335.

Zhang, J., Hsee, C. K., & Xiao, Z. (2006). The majority rule in individual decision making. *Organizational Behavior and Human Decision Processes, 99*, 102-111.

Table 1. Four classes of models and their implications for transitivity and restricted branch independence. The majority rule satisfies restricted branch independence and violates transitivity; the configural weight model violates restricted branch independence and satisfies transitivity.

Restricted branch independence	Transitivity	
	Satisfied (Figure 3)	Violated (Figure 4)
Satisfied	Constant weight averaging, expected utility, subjectively weighted utility. (Equation 4).	Additive contrast model (majority rule, regret theory, stochastic difference, PRAM, salience weighted utility, similarity models. (Equations 1 and 10).
Violated	Configural weight averaging model (includes cumulative prospect theory, TAX model). (Equation 5).	Lexicographic semi-orders, priority heuristic. See Birnbaum (2010).

Table 2. Estimation of error rates in tests of transitivity (Experiments 1 and 2). Chi-Squares show fit of the true and error (TE) model and the fit of response independence (Indep) to the same data, with same number of parameters.

Choice (XY)	Choice		Response Pattern				Choice %*Y*	Parameter estimates		Statistical Tests	
	X	*Y*	*XX**	*XY**	*YX**	*YY**		*p*	*e*	$\chi^2(1)$ TE	$\chi^2(1)$ Indep
Experiment 1											
FG	(4, 5, 5)	(2, 6, 7)	62	25	24	103	0.60	0.63	0.13	0.02	58.90
GH	(2, 6, 7)	(3, 7, 3)	163	18	13	20	0.17	0.10	0.08	0.80	49.05
HF	(3, 7, 3)	(4, 5, 5)	10	9	12	183	0.90	0.95	0.05	0.43	40.55
Experiment 2											
FG	(4, 5, 6)	(5, 7, 3)	84	41	54	132	0.58	0.62	0.19	1.77	44.12
GH	(5, 7, 3)	(9, 1, 5)	156	40	35	82	0.38	0.34	0.14	0.33	76.02
HF	(9, 1, 5)	(4, 5, 6)	61	31	28	191	0.71	0.77	0.11	0.15	90.84
F'G'	(6, 4, 5)	(5, 7, 3)	172	30	49	62	0.32	0.25	0.15	4.52	58.04
G'H'	(5, 7, 3)	(1, 5, 9)	208	31	24	49	0.25	0.18	0.10	0.89	86.01
H'F'	(1, 5, 9)	(6, 4, 5)	77	17	26	193	0.69	0.72	0.08	1.87	146.15

Table 3. Analysis of response patterns in test of transitivity in Experiment 1. Rep = replicate.

Choice Pattern			Observed Frequency			Estimated probabilities
FG	*GH*	*HF*	Rep 1	Rep 2	Both	
F	*G*	*H*	4	4	1	0.01
F	*G*	*F*	62	59	30	0.27
F	*H*	*H*	3	2	1	0.01
F	*H*	*F*	18	21	9	0.08
G	*G*	*H*	11	15	5	0.03
G	*G*	*F*	104	98	65	0.60
G	*H*	*H*	1	1	0	0.00
G	***H***	***F***	11	14	2	0.01
Totals			214	214	113	1

Table 4. Tests of restricted branch independence in Experiment 1. % *R* = percentage choosing the risky gamble.

	Choice		
Type	Safe (*S*)	Risky (*R*)	%*R*
Design 1			
SR	(1, 4, 6)	(1, 2, 9)	37
S'R'	(5, 4, 6)	(5, 2, 9)	53
S"R"	(10, 4, 6)	(10, 2, 9)	66
Design 2			
SR	(2, 4, 5)	(2, 2, 7)	19
S"R"	(9, 4, 5)	(9, 2, 7)	48

Table 5. Analysis of response patterns in tests of transitivity and recycling (Experiment 2). Totals are less than the sample size because a few participants skipped at least one of the six choice problems analyzed in a design. Patterns *FGH* and *GHF* are intransitive. *p* are the estimated true probabilities of the response patterns.

	FGH design				*F′ G′ H′* design			
Pattern	Rep 1	Rep 2	Both	Est *p*	Rep 1	Rep 1	Both	Est *p*
FGH	11	13	2	0.01	41	45	29	0.15
FGF	90	91	58	0.39	130	129	90	0.53
FHH	8	17	2	0.00	9	19	6	0.03
FHF	14	15	1	0.00	18	24	5	0.02
GGH	28	20	4	0.04	14	14	2	0.01
GGF	63	63	25	0.18	51	41	19	0.12
GHH	45	39	24	0.19	29	24	13	0.11
GHF	47	48	28	0.18	17	13	3	0.02
	306	306	144	1	309	309	167	1

Table 6. Choice types, prize values, and predictions of regret theory (Regret), majority rule (MR) and transfer of attention exchange (TAX) models for Experiments 3 and 4. TAX model predictions are based on prior parameters. Regret predictions are based on regret functions estimated by Bleichrodt, et al. (2010). Values of prizes in Experiment 4 were one-tenth those in Experiment 3 ($10 to $90). The predicted response pattern for regret is denoted *111* (choice of the first listed alternative in all three choices, *A*, *B*, and *C* in *AB*, *BC*, and *CA* choices, respectively); for majority rule (advantage seeking) is *222*, and for TAX, it is *112*.

Choice Problem			Theories		
Type	First	Second	Regret	MR	TAX
AB	*A* = (400, 500, 600)	*B* = (500, 700, 300)	*A*	*B*	*A*
BC	*B* = (500, 700, 300)	*C* = (900, 100, 500)	*B*	*C*	*B*
CA	*C* = (900, 100, 500)	*A* = (400, 500, 600)	*C*	*A*	*A*
A'B'	*A'* = (600, 400, 500)	*B'* = (500, 700, 300)	*B'*	*A'*	*A'*
B'C'	*B'* = (500, 700, 300)	*C'* = (100, 500, 900)	*C'*	*B'*	*B'*
C'A'	*C'* = (100, 500, 900)	*A'* = (600, 400, 500)	*A'*	*C'*	*A'*

Table 7. Number of response patterns in the first and second replicates of the *ABC* design, summed over blocks (Experiment 3). The patterns, *111* and *222* are the intransitive response patterns that are predicted by regret theory and majority rule, respectively. Rep = replicate.

Response Pattern First Rep	Response Pattern on Second Replicate *ABC*							
	111	*112*	*121*	*122*	*211*	*212*	*221*	*222*
111	2	7	1	1	3	2	1	0
112	6	101	3	8	3	24	3	6
121	1	0	5	1	2	1	9	2
122	3	7	7	9	6	1	4	3
211	1	6	3	1	3	2	3	4
212	6	43	1	5	9	32	1	6
221	4	1	9	2	11	3	32	5
222	1	10	1	8	5	7	5	18

Table 8. Number of response patterns in the *A'B'C'* design, as in Table 7 (Experiment 3). Regret theory implies the intransitive pattern *222* and majority rule predicts *111*, opposite of the predictions for *ABC* design of Table 7.

Response Pattern First Rep	Response Pattern on Second Replicate *A'B'C'*							
	111	*112*	*121*	*122*	*211*	*212*	*221*	*222*
111	14	12	4	4	2	1	1	3
112	10	143	6	10	8	23	1	3
121	5	3	12	3	2	2	6	3
122	1	10	6	4	3	1	4	4
211	2	1	1	2	4	3	4	4
212	3	34	5	8	4	16	3	8
221	2	2	6	5	4	3	16	1
222	1	4	2	4	6	3	0	5

Table 9. Number of people who showed each combination of preferences in *ABC* and *A'B'C'* designs, summed over both repetitions and blocks. The intransitive, recycling pattern *222111*, predicted by majority rule, was observed a total of 27 times. The opposite pattern, *111222*, predicted by regret theory, was observed 9 times. (Experiment 3)

Response Pattern on *ABC*	Response Pattern on *A'B'C'*							
	111	*112*	*121*	*122*	*211*	*212*	*221*	*222*
111	4	12	2	5	2	4	3	9
112	8	238	3	12	6	51	3	8
121	8	11	9	8	8	1	5	1
122	7	20	14	7	5	6	11	5
211	5	13	6	10	2	13	7	9
212	13	89	8	7	6	40	3	9
221	7	12	25	15	15	8	35	8
222	27	18	11	9	10	10	7	7

Table 10. Parameter estimates and index of fit for true and error models fit to tests of transitivity and recycling in Experiments 3 and 4. Values in parentheses are fixed or constrained. MR = majority rule.

Parameter	Experiment 3			Experiment 4		
	General	MR	Transitive	General	MR	Transitive
p_{111}	0.00	(0)	(0)	0.00	(0)	(0)
p_{112}	0.57	0.56	0.59	0.65	0.64	0.67
p_{121}	0.05	0.04	0.04	0.01	0.02	0.02
p_{122}	0.02	0.03	0.05	0.02	0.02	0.03
p_{211}	0.01	0.01	0.01	0.02	0.02	0.02
p_{212}	0.10	0.12	0.12	0.12	0.13	0.13
p_{221}	0.15	0.16	0.18	0.11	0.12	0.13
p_{222}	0.02	(0)	(0)	0.01	(0)	(0)
p_R	0.01	(0)	(0)	0.01	(0)	(0)
p_{MR}	0.07	0.08	(0)	0.06	0.06	(0)
e_1	0.27	0.23	0.25	0.25	0.20	0.21
e_2	0.13	0.14	0.16	0.12	0.09	0.10
e_3	0.13	0.14	0.16	0.10	0.09	0.10
e'_1	0.19	(=e_1)	(=e_1)	0.16	(=e_1)	(=e_1)
e'_2	0.14	(=e_2)	(=e_2)	0.08	(=e_2)	(=e_2)
e'_3	0.13	(=e_3)	(=e_3)	0.06	(=e_3)	(=e_3)
f_1	--	--	--	0.19	(=e_1)	(=e_1)
f_2	--	--	--	0.08	(=e_2)	(=e_2)
f_3	--	--	--	0.09	(=e_3)	(=e_3)
ΣG^2	178.5	190.8	248.4	363.7	404.3	527.6

Table 11. Observed proportion of choices for second gamble in tests of restricted branch independence and other "filler" choice problems (Experiments 3 and 4). The consequences in Experiment 4 were one-tenth of the values listed below ($10 to $90). RBI = restricted branch independence; MR-EV = majority rule versus expected value; DOM = transparent dominance.

Choice Problem			Choice Proportion	
Type	First	Second	Experiment 3	Experiment 4
RBI1	(100, 400, 600)	(100, 100, 900)	0.17	0.16
RBI1	(500, 400, 600)	(500, 100, 900)	0.31	0.20
RBI1	(900, 400, 600)	(900, 100, 900)	0.48	0.40
RBI2	(100, 400, 500)	(100, 100, 900)	0.21	0.29
RBI2	(900, 400, 500)	(900, 100, 900)	0.53	0.56
MR-EV1	(500, 500, 500)	(600, 100, 600)	0.15	0.14
MR-EV2	(200, 300, 500)	(600, 200, 400)	0.83	0.87
DOM1	(200, 300, 500)	(200, 400, 600)	0.89	0.95
DOM2	(200, 200, 800)	(200, 800, 800)	0.92	0.97
DOM3	(400, 500, 900)	(400, 400, 800)	0.12	0.09

Table 12. Percentage of repeated (consistent) patterns in tests of transitivity and recycling (Exp = Experiments 4, 5 and 6). Last row shows total number of participant-blocks (which may reflect occasional skipped items); next to last row shows the percentage of consistent choice patterns (response patterns were the same in both repetitions within blocks). Here I = (90, 50, 10), J = (10, 90, 50), K = (50, 10, 90). See Tables SM.3-SM.12 for details.

	Exp 4	Exp 4	Exp 4	Exp 6	Exp 6	Exp 6	Exp 5	Exp 6	Exp 6
Pattern	*ABC*	*A'B'C'*	*A"B"C"*	*ABC*	*A'B'C'*	*A"B"C"*	*IJK*	*IJK*	*I'J'K'*
111	1	5	0	2	0	2	5	10	0
112	59	72	71	62	62	53	6	12	22
121	2	1	1	0	3	5	9	2	5
122	5	1	1	0	3	0	39	40	5
211	2	3	1	0	0	0	5	12	46
212	17	12	14	13	12	13	1	2	5
221	9	7	10	23	18	27	16	19	8
222	5	0	1	0	2	0	18	2	8
% repeats	48	56	52	49	60	60	33	42	37
Total No.	405	405	405	96	100	100	1118	100	100

Table 13. Estimated true probabilities of choosing the risky gamble (p) and their error rates (e) in four tests of restricted branch independence. (Experiments 5 and 6). First and Last represent the first and last blocks of trials from Experiment 5, respectively. See Tables SM.13-SM.15 for details.

		Choice		Experiment 5, First		Experiment 5, Last		Experiment 6	
Test	Type	Safe, *S*	Risky, *R*	*p*	*e*	*p*	*e*	*p*	*e*
4	*SR*	(5, 30, 40)	(5, 10, 90)	0.54	0.10	0.64	0.08	0.32	0.07
4	*S'R'*	(95, 30, 40)	(95, 10, 90)	0.84	0.12	0.79	0.05	0.61	0.08
5	*SR*	(5, 35, 45)	(5, 10, 90)	0.31	0.25	0.52	0.06	0.27	0.05
5	*S'R'*	(95, 35, 45)	(95, 10, 90)	0.71	0.07	0.75	0.08	0.46	0.09
6	*SR*	(5, 40, 50)	(5, 10, 90)	0.39	0.13	0.44	0.10	0.19	0.07
6	*S'R'*	(95, 40, 50)	(95, 10, 90)	0.68	0.16	0.62	0.07	0.38	0.02
7	*SR*	(5, 45, 55)	(5, 10, 90)	0.23	0.08	0.37	0.07	0.14	0.05
7	*S'R'*	(95, 45, 55)	(95, 10, 90)	0.52	0.21	0.56	0.11	0.25	0.10

Figure 1. Format of one trial from Experiments 1 and 2. The participant's task was to decide which city to visit based on advice from three friends who rated how much they think the participant would like to visit those cities, using a scale in which 10 is the best and 1 is the worst.

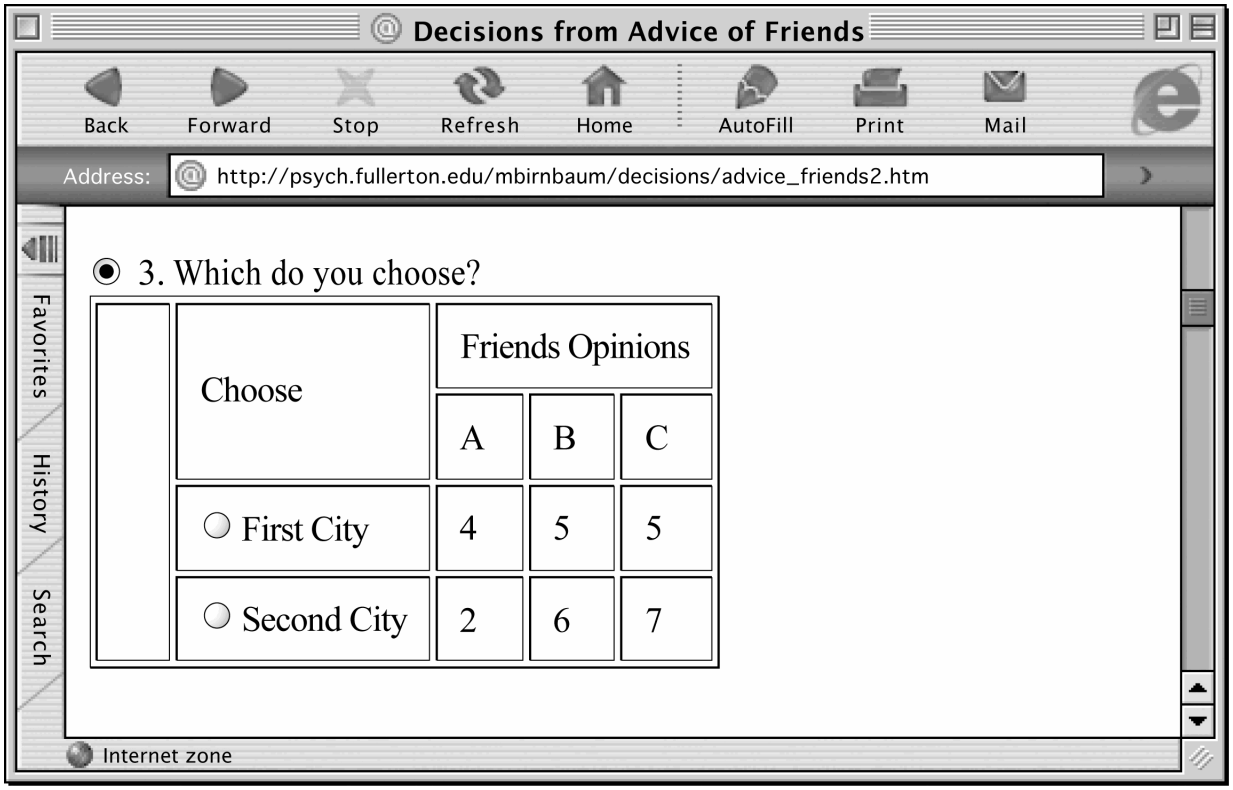

Figure 2. Display of a choice problem as in Experiments 4 and 5. The participant's task was to choose which gamble to play, based on the prizes to be won if a Red, White, or Blue marble would be drawn at random from an urn containing an equal number of each color.

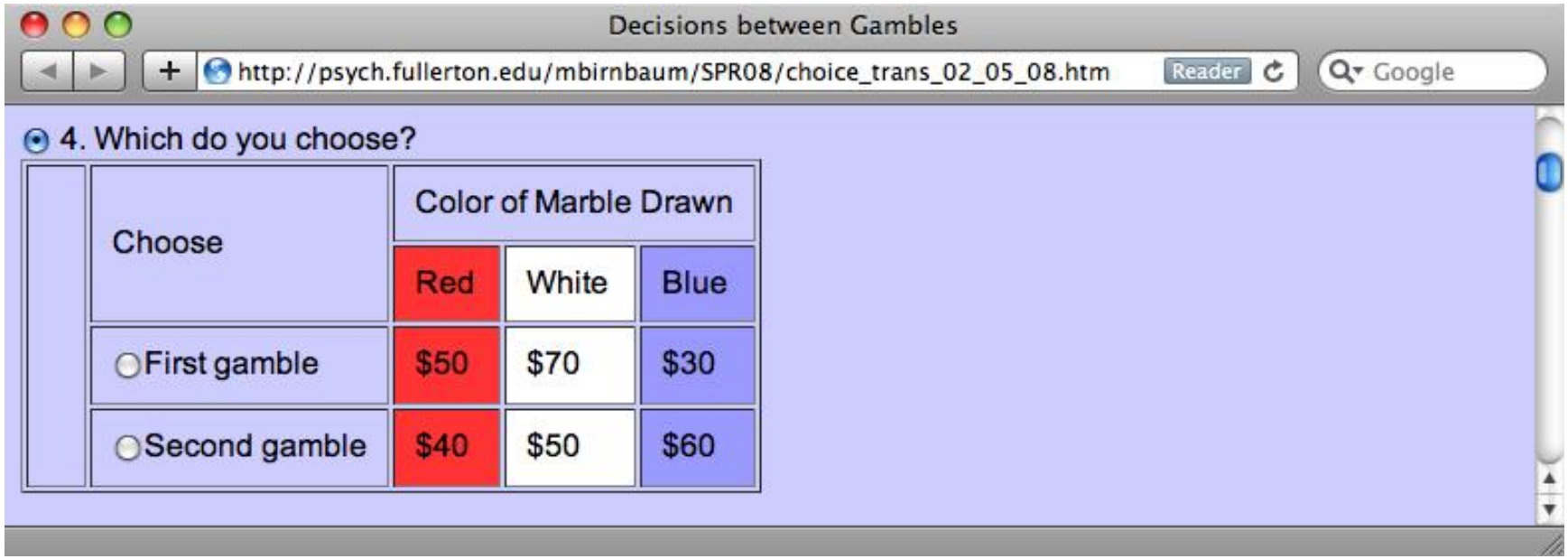

Figure 3. Outline of a transitive model of the comparison of two cities or of two gambles. *F* and *G* are alternatives of a choice problem; *x*, *y*, and *z* represent either the ratings of cites by three friends or monetary prizes of gambles depending on whether a Red, White, or Blue marble is drawn from an urn. The objective values, *x*, *y*, and *z* are transformed to subjective values, $u(x)$, $u(y)$, and $u(z)$ by the evaluation process (*V*); these are integrated to compute overall evaluations of the alternatives, $U(F)$ and $U(G)$, by the integration function (*I*), and compared via *C* to produce a subjective comparison, δ, which is mapped to an overall response by the judgment function, *J*. The small arrows pointing southwest indicate loci where random errors might enter the process.

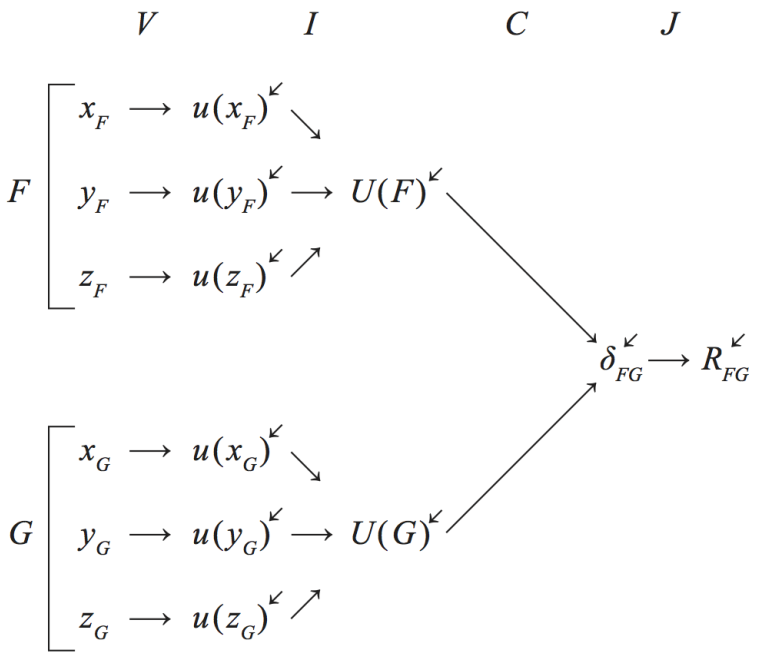

Figure 4. Outline of an intransitive choice process, as in Figure 3. In this case, evaluations of the components of the alternatives are compared prior to integration. The contrasts, ψ, represent relative arguments for one alternative or the other based on comparisons of the evaluations of the components. These contrasts are then integrated to form an overall preference, δ, which is mapped to an overt response.

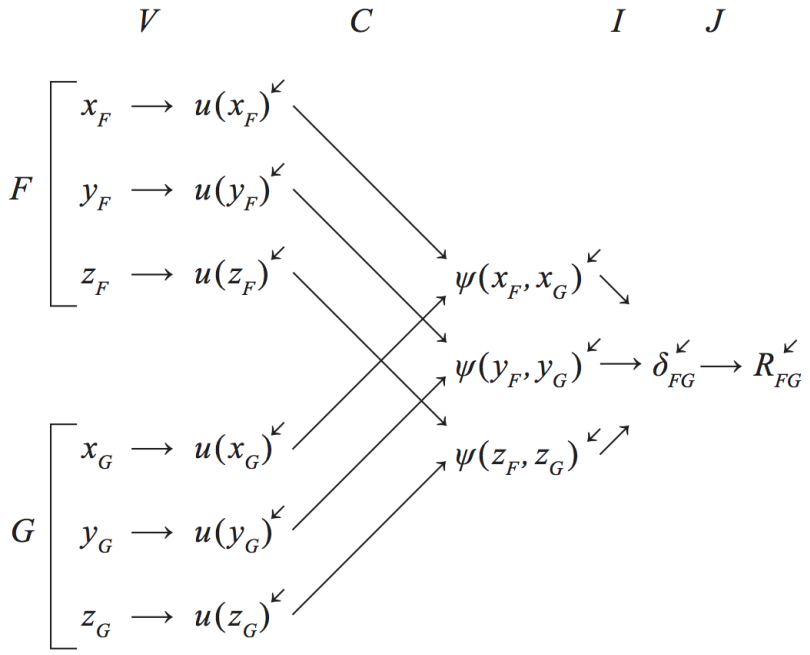

Figure 5. Predicted cycles and reversed cycles in the *ABC* and *A'B'C'* designs. Arrowheads represent preference relations. For example, Majority rule (outer, dashed curves on the left) predicts the intransitive cycle, $A \succ C \succ B \succ A$. Note that regret and majority rule produce opposite intransitive cycles, but both produce recycling.

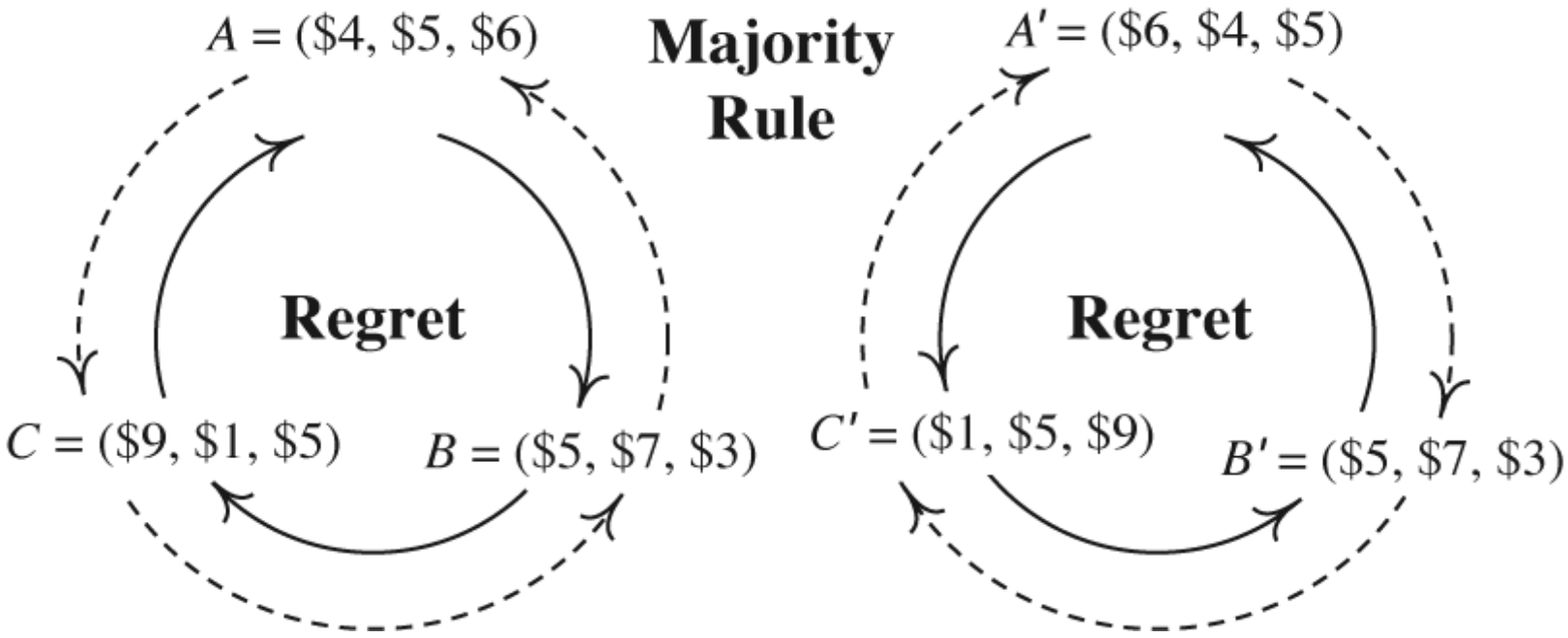

Figure 6. Theoretical analysis of *ABC* and *A'B'C'* designs in Experiment 4 as related to parameters α and β of Equation 13. Bleichrodt, et al. estimated α = 1 and β slightly greater than 1.5 (strong regret aversion). With these parameters, Equation 13 predicts *111* intransitive cycle in the *ABC* design and the opposite intransitive cycle, *222*, in the *A'B'C'* design.