Cognitive Psychology 118 (2020) 101258

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Modeling choice paradoxes under risk: From prospect theories to sampling-based accounts☆

David Kellen[a,*], Markus D. Steiner[b], Clintin P. Davis-Stober[c], Nicholas R. Pappas[a]

[a] Syracuse University, United States
[b] University of Basel, Switzerland
[c] University of Missouri, United States

ARTICLE INFO

Keywords:
Choices under risk
Utility
Prospect theory
Choice paradoxes
Sequential sampling

ABSTRACT

Important developments in the study of decision making have been based on the establishment and testing of choice paradoxes (e.g., *Allais'*) that reject different theories (e.g., Expected Utility Theory). One of the most popular and celebrated models in the literature, *Cumulative Prospect Theory* (CPT), has managed to retain its status despite a growing body of empirical evidence stemming from a collection of choice paradoxes that reject it. Two alternative models, *Transfer of Attention Exchange* (TAX) and an extension of *Decision Field Theory* (DFT_e), have been proposed as possible alternatives to CPT. To date, no study has directly compared these three models within the context of a large set of lottery problems that tests different choice paradoxes. The present study accomplishes this by using a large and diverse set of lottery problems, involving both potential gains and losses. Our results support the presence and robustness of a set of 'strong' choice paradoxes that reject CPT irrespective of its parametric form. Model comparison results show that DFT_e provides the best account for the present set of lottery problems, as it is able to accommodate the choice data at large in a parsimonious fashion. The success of DFT_e shows that many behavioral phenomena, including paradoxes that CPT cannot account for, can be successfully captured by a simple noisy-sampling process. Overall, our results suggest that researchers should move away from CPT, and focus their efforts on alternative models such as DFT_e.

The study of decision making under risk is concerned with individual preferences between options that yield outcomes with known probabilities. The focal point of this research is to understand how desire and possibility are represented by decision makers and ultimately integrated into a single value that allows for comparisons to be made and preferences to be expressed. For example, consider the comparison between an option $\mathcal{A} = \begin{pmatrix} \$100 \\ 1 \end{pmatrix}$ that always yields \$100 for sure, and a lottery $\mathcal{B} = \begin{pmatrix} \$200 & \$0 \\ .50 & .50 \end{pmatrix}$ that establishes two equiprobable mutually-exclusive events, one yielding \$200, and the other \$0. Whereas one option consists of a sure monetary gain, the other option involves a better monetary gain that is only expected to occur half of the time (with nothing being received otherwise). The preference for one option over the other, let us say $\mathcal{A}$ over $\mathcal{B}$, is expressed by $\mathcal{A} \succ \mathcal{B}$, using the binary operator '$\succ$' (operator $\prec$ would express the opposite preference).

Theoretical work in decision making under risk has traditionally relied on choice paradoxes (e.g., *St Petersburg's*, *Allais'*) to reject large classes of models. The most prominent model in the literature, *Cumulative Prospect Theory* (CPT), has managed to retain this

☆ Author Note: David Kellen received support from the Swiss National Science Foundation Grant 100014_165591.

* Corresponding author at: Department of Psychology, Syracuse University, United States.
E-mail address: davekellen@gmail.com (D. Kellen).

https://doi.org/10.1016/j.cogpsych.2019.101258
Received 11 February 2018; Received in revised form 23 October 2019; Accepted 26 November 2019
Available online 10 February 2020
0010-0285/

status despite a growing body of empirical evidence on choice paradoxes that reject it. Two alternative models, *Transfer of Attention Exchange* (TAX) and an extension of *Decision Field Theory* (DFT_e), have been proposed as possible alternatives to CPT. Yet, no direct attempt has been to compare the three models in the context of these choice paradoxes. The present study accomplishes this by using a large and diverse set of lottery problems, involving potential gains and losses, based upon a collection of choice paradoxes that challenge CPT. Results show strong empirical evidence for the choice paradoxes, demonstrating their robustness in this experimental setting. Model comparison results show that DFT_e provides the best account, as it is able to accommodate the data at large (including the aforementioned new choice paradoxes) in a rather parsimonious fashion. Overall, results suggest that researchers should move away from CPT, towards sequential-sampling models such as DFT_e.

Theoretical developments in this domain have largely been dominated by '*choice paradoxes*' — choice patterns that are at odds with the core tenets of a given theory. For instance, a theory might assume that a certain option $\mathcal{D}$ can never be preferred to another option $\mathcal{C}$, or that the preference in a given pair necessarily implies a certain preference in a closely related pair: For example, $\mathcal{C} \succ \mathcal{D}$ implies $\mathcal{C}' \succ \mathcal{D}'$. In the first case, a choice paradox would then consist of reliably observing $\mathcal{D} \succ \mathcal{C}$. In the second case, it would consist of reliably observing both $\mathcal{C} \succ \mathcal{D}$ and $\mathcal{D}' \succ \mathcal{C}'$. Because they enable the rejection of a large classes of theories based on a critical set of observations, choice paradoxes can be seen as a form of strong inference (Platt, 1964).

Consider lottery $\mathcal{L} = \begin{pmatrix} x_1 & x_2 & \dots & x_K \\ p_1 & p_2 & \dots & p_K \end{pmatrix}$, comprised of K mutually-exclusive events, each occurring with probability p_k and yielding an outcome x_k. [1] According to the *Expected Utility Theory* (EUT; von Neumann & Morgenstern, 1944), the subjective value of a lottery corresponds to a probability-weighted sum of the subjective values of each monetary outcome:

$$V(\mathcal{L}) = \sum_{k=1}^{K} u(x_k)\cdot p_k. \tag{1}$$

According to EUT, the subjective representation of monetary values is given by a function $u(x)$. One important aspect of EU is its ability to account for the famous *St. Petersburg's Paradox* (Bernoulli, 1738/1954), in which individuals find a game with infinite expected value to be unattractive. This behavioral 'paradox' is attributed to the notion that the subjective value of constant increments gets smaller and smaller (e.g., $u(\$120) - u(\$110) \leqslant u(\$20) - u(\$10)$), which can be captured by a concave function $u(x)$. But despite its initial successes, EU was later shown to be unable to predict a set of preference patterns known as the *Allais' Paradoxes* (Allais, 1953). For instance, consider Allais' *Common Consequence Paradox*, demonstrated by the two lottery problems below:

$$\mathcal{A} = \begin{pmatrix} \$2400 \\ 1 \end{pmatrix} \quad \mathcal{B} = \begin{pmatrix} \$2500 & \$2400 & \$0 \\ .33 & .66 & .01 \end{pmatrix},$$

$$\mathcal{A}' = \begin{pmatrix} \$2400 & \$0 \\ .34 & .66 \end{pmatrix} \quad \mathcal{B}' = \begin{pmatrix} \$2500 & \$0 \\ .33 & .67 \end{pmatrix}.$$

According to EUT, a preference for $\mathcal{A}/\mathcal{A}'$ necessarily implies a preference for $\mathcal{B}/\mathcal{B}'$. To understand this prediction note that $\mathcal{A}$ is equivalent to $\begin{pmatrix} \$2400 & \$2400 \\ .34 & .66 \end{pmatrix}$, which means that the difference between the two lottery problems amounts to a change of the outcome (from \$2400 to \$0) of a common event with probability .66. EUT predicts that such changes cannot lead to a preference reversal, but as shown by Kahneman and Tversky (1979), 82% of respondents preferred $\mathcal{A}$ over $\mathcal{B}$, but 83% preferred $\mathcal{B}'$ over $\mathcal{A}'$. The *Allais' Common Ratio Paradox* consists of a similar preference reversal:

$$\mathcal{A} = \begin{pmatrix} \$3000 \\ 1 \end{pmatrix} \quad \mathcal{B} = \begin{pmatrix} \$4000 & \$0 \\ .80 & .20 \end{pmatrix},$$

$$\mathcal{A}' = \begin{pmatrix} \$3000 & \$0 \\ .25 & .75 \end{pmatrix} \quad \mathcal{B}' = \begin{pmatrix} \$4000 & \$0 \\ .20 & .80 \end{pmatrix}.$$

EUT again expects an invariance such that a preference for $\mathcal{A}$ over $\mathcal{B}$ should only happen if there is a preference for $\mathcal{A}'$ over $\mathcal{B}'$. The reason is that the outcomes are the same and the probabilities have the same ratio (i.e.,$1/.8 = .25/.20$). Kahenman and Tversky (1979) reported that 80% of the respondents preferred $\mathcal{A}$ but 65% preferred $\mathcal{B}'$.

Phenomena such as the Allais' paradoxes led to the development of theoretical accounts such as *Cumulative Prospect Theory* (CPT), which assumes a *subjective representation of probabilities* and a weighting of outcomes that depends on their sign (gain vs. loss) and/or rank (Tversky & Kahneman, 1992). A major point of contention between the different theories – among which CPT is the most popular – is the exact way in which outcomes are weighted (for reviews, see Fehr-Duda & Epper, 2012; Luce, 2000; Wakker, 2010). These differences are not immaterial, given that the different subjective representations postulated by these models are often used to characterize individual and group differences (e.g., Booij & Van de Kuilen, 2009; Kellen, Pachur, & Hertwig, 2016; Pachur, Mata, & Hertwig, 2017; Patalano, Saltiel, Machlin, & Barth, 2015; Schley & Peters, 2014).

In this manuscript, we revisit the '*new choice paradoxes*' developed by Birnbaum and colleagues, which are directed at falsifying models such as CPT (for a review, see Birnbaum, 2008). We report the results of a new study designed to accomplish two major goals:

[1] We will refer to lotteries with two events as *binary lotteries*, and three or more events as *multiple-event lotteries*. Also, we will refer to any lottery comprised exclusively of events with non-negative/non-positive outcomes as *pure-gain/pure-loss* lotteries. Lotteries with potential gains and losses are referred to as *mixed*. Finally, we will refer to any pair of lotteries an individual has to choose from as a *lottery problem*.

First, assess the robustness of these choice paradoxes, examining their replicability. Second, provide a direct empirical comparison between three candidate models, namely Cumulative Prospect Theory (CPT; Tversky & Kahneman, 1992), the *Transfer of Attention Exchange* model (TAX; Birnbaum, 2008), and an extended version of the (simplified) *Decision Field Theory* model (DFT_e; Bhatia, 2014; Busemeyer & Townsend, 1992). Our study is designed to not solely focus on these choice paradoxes, but to embed them within a larger set of stimuli designed to test CPT-related phenomena, including lottery problems sensitive to so-called strong risk attitudes and gain-loss asymmetries. The end goal is to provide a critical, yet holistic test of these general properties and an appropriate testbed for evaluating the three major models under consideration. To date, these three models have not been subjected to a direct comparison in which their performance is evaluated at the individual-participant level using a large and diverse set of lottery problems.

The manuscript is organized as follows: we will first describe all three theories in detail, as well as the choice paradoxes originally used to compare CPT and TAX. In a preliminary step, we will fit all models to individual-participants' choices coming from a study by Rieskamp (2008). We will then discuss the present experiment and evaluate the occurrence of different choice patterns, including the aforementioned choice paradoxes, and report a penalized-fit comparison of the three theories.

1. Cumulative prospect theory

According to CPT, individuals' preferences are a function of subjective representations of monetary outcomes x associated to each possible event, and the probabilities p of each event taking place. Consider a lottery $\mathcal{L} = \begin{pmatrix} x_1^- & \dots & x_M^- & x_1^+ & \dots & x_N^+ \\ p_1^- & \dots & p_M^- & p_1^+ & \dots & p_N^+ \end{pmatrix}$. In CPT, lottery branches (i.e., a lottery's event-probability pairs) are ordered according to the outcomes' sign and rank: For loss outcomes x^-, outcomes are ordered from worst to best, $x_1^- \leqslant \dots \leqslant x_M^-$. In contrast, gain outcomes x^+ are ordered from best to worst, $x_1^+ \geqslant \dots \geqslant x_N^+$. The subjective value V of lottery $\mathcal{L}$ is given by

$$V(\mathcal{L}) = \sum_{m=1}^{M} u(x_m^-)W_m^- + \sum_{n=1}^{N} u(x_n^+)W_n^+, \tag{2}$$

where $u(x)$ is the *utility function* and W the *decision-weight function*. For each outcome x there is a decision weight W that consists of the subjective probability of receiving at least x *minus* the subjective probability of receiving any of the lottery's same-domain outcomes that ranks above x, i.e., a difference between weighted cumulative probabilities (Quiggin, 1982). In CPT, decision weights are computed separately for gains and losses. Gains are ranked from best to worst, and losses from worst to best. Decision weights for gains and losses are computed as follows:

$$W_k = w\left(\sum_{k' \leqslant k} p_{k'}\right) - w\left(\sum_{k'' < k} p_{k''}\right), \tag{3}$$

where subscripts k, k', and k'' denote ranks.

CPT's utility function attempts to capture the subjective representation of monetary gains and losses. The following parametric function is commonly used:

$$u(x) = \begin{cases} x^\alpha, & \text{for } x \geqslant 0, \\ -\lambda|x|^\beta, & \text{for } x < 0. \end{cases} \tag{4}$$

As shown in the left panel of Fig. 1, when parameters α/β are between 0 and 1, function $u(x)$ is concave for gains and convex for losses, capturing the notion that absolute increments in gains and losses lead to diminishing changes in subjective value. Also, when parameter λ is larger than 1, subjective values associated with losses (e.g., lose \$1) tend to be perceived as greater than their gain

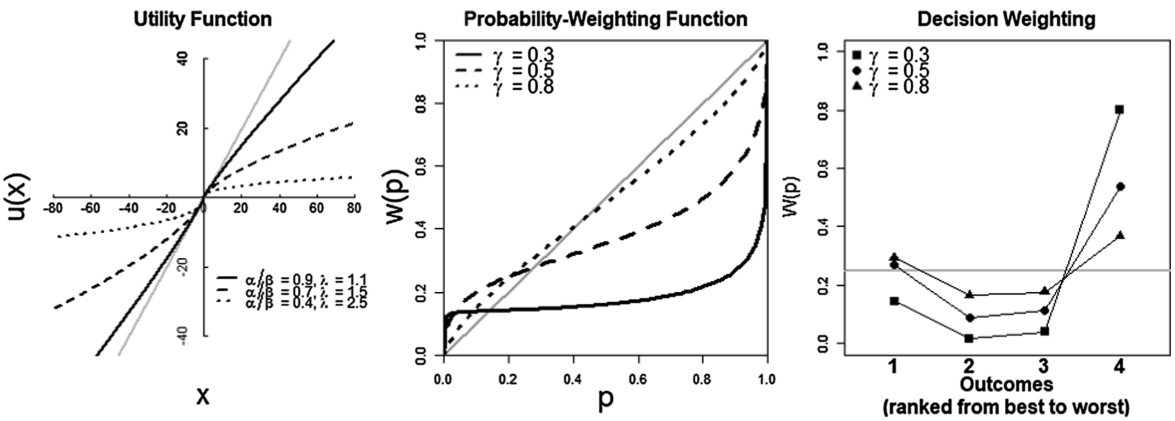

Fig. 1. Illustration of CPT's utility function (left panel), probability weighting function (center panel), and decision weighting function (right panel).

counterparts (e.g., gain $1), i.e., loss aversion (Köbberling & Wakker, 2005).[2] Parameter λ attempts to capture behavioral gain-loss asymmetries, such as participants' general dislike for lotteries with symmetrical, equiprobable outcomes.

For example $\begin{pmatrix} x & -x \\ .50 & .50 \end{pmatrix}$, would be seen as increasingly less attractive as x becomes more extreme. This means that loss-averse individuals are expected to prefer $\mathcal{A} = \begin{pmatrix} \$10 & -\$10 \\ .50 & .50 \end{pmatrix}$ to $\mathcal{B} = \begin{pmatrix} \$15 & -\$15 \\ .50 & .50 \end{pmatrix}$, for example.[3] When $\alpha, \beta \leqslant 1$, the utility function $u(x)$ is concave for gains and convex for losses. The shape of $u(x)$ yields specific predictions in terms of so-called *strong risk attitudes* under a range of probabilities $[d, 1]$ for which the probability-weighting function $w(p)$ (discussed below) is convex (e.g., $d \geqslant .20$ under typical parameter values; for details, see Baucells & Heukamp, 2006). Consider lottery $\mathcal{A} = \begin{pmatrix} x & y \\ .50 & .50 \end{pmatrix}$, and the mean-preserving spread $\mathcal{B} = \begin{pmatrix} x+z & y-z \\ .50 & .50 \end{pmatrix}$. For pure-gain lotteries with $x > y > z > 0$, individuals are expected to prefer $\mathcal{A}$ to $\mathcal{B}$ (be risk averse). But in the case of loss-only lotteries (with $x < y < z < 0$), they are expected to prefer $\mathcal{B}$ to $\mathcal{A}$ (be risk seeking). Empirical results consistent with these functional shapes were reported by Brooks, Peters, and Zank (2014).

CPT assumes a subjective representation of probabilities that deviates from their objective counterparts in a systematic way (for an early proposal, see Edwards, 1954, Edwards, 1962). As reviewed by Kahneman and Tversky (1979), individuals' preferences tend to be very sensitive to changes in extreme probabilities (e.g., going from $p = 1$ to $p = .99$) and relatively insensitive to differences between moderate probabilities (e.g., going from $p = .45$ to $p = .55$). These patterns can be captured by an inverse-S-shaped representation of probabilities $w(p)$, like the one illustrated in Fig. 1. CPT accommodates the Allais' paradoxes, such as the Common-Consequence Paradox, via its decision weights and a non-linear probability representation that overweights small probability values (e.g., event $0 in $\mathcal{B}$). According to Kahneman and Tversky, inverse-S-shaped probability-weighting functions can be approximated by the following parametric form:

$$w(p) = \frac{p^{\gamma}}{(p^{\gamma} + (1-p)^{\gamma})^{\frac{1}{\gamma}}}, \tag{5}$$

where γ is a probability sensitivity parameter that yields an inverse-S-shaped function for values between 0.28 and 1 and an S-shaped function for values larger than 1.[4] Note that when $\gamma = 1$, $w(p) = p$. Alternative parametric forms have been proposed since then (e.g., Goldstein & Einhorn, 1987; Prelec, 1998). The rank-dependent decision weights W obtained with an inverse-S-shaped function $w(p)$ yield a mixed pattern of pessimism/optimism, with individuals placing large weights in the best and worst outcomes as shown in the right panel Fig. 1, where we illustrate the weights for a lottery with four equiprobable events.

Finally, let us discuss how differences between the lotteries' values can be translated into choice probabilities. One common approach for capturing the stochastic nature of choices is to implement a logistic 'choice rule' (Luce, 1953; for an application, see Rieskamp, 2008):

$$P(\mathcal{A} \succ \mathcal{B}) = \frac{1}{1 + \exp(-\theta(V(\mathcal{A}) - V(\mathcal{B}))}. \tag{6}$$

Choice probabilities become more extreme (i.e., close to 0 or 1) as the difference between the values of two lotteries increases. The choice-sensitivity parameter θ modulates how much choice probabilities change as a function of the differences between the options' subjective values.

2. New choice paradoxes

CPT has been subjected to intensive empirical scrutiny, most notably by Birnbaum and colleagues, who developed a large set of *new choice paradoxes* that reject it (for a review, see Birnbaum, 2008). These choice paradoxes involve lottery problems constructed using numerous 'recipes' that rely on some of the assumptions underlying CPT (e.g., coalescing, monotonicity, transitivity, comonotonic independence). The paradoxes discussed here fall into one of three scenarios: In the first scenario, there is a lottery pair that includes an option that, according to CPT, *cannot* be preferred (unless some error occurs). The second scenario involves a set of two lottery problems involving two pairs of options, $\{\mathcal{A}, \mathcal{B}\}$ and $\{\mathcal{A}', \mathcal{B}'\}$. According to CPT, a preference for $\mathcal{A}/\mathcal{B}$ necessarily implies a preference for $\mathcal{A}'/\mathcal{B}'$. The Allais' paradoxes fall into this second category. The third scenario is one in which the preference for a specific option in a lottery problem (e.g., $\mathcal{B}$) implies a preference for another option in a second lottery problem ($\mathcal{B}''$). The violation of any of these preference constraints are considered to be 'paradoxical' given that they imply the rejection of at least some of the

[2] As discussed in detail by Nilsson, Rieskamp, and Wagenmakers (2011), it is extremely difficult to jointly estimate β and λ. The problem is that they can end up serving very similar roles, in addition to the fact that λ can only be estimated in the context of mixed lotteries. This situation has led many to impose the constraint $\alpha = \beta$ (e.g., Glöckner & Pachur, 2012; Kellen et al., 2016). This equality enforces a shape constraint, with a concave/convex value function for gains implying a convex/concave function for losses. In light of recent results questioning the appropriateness of this constraint (e.g., Abdellaoui, Bleichrodt, & L'Haridon, 2008; Davis-Stober and Brown, 2013; Kellen, Mata, & Davis-Stober, 2017), we will somewhat relax it later on in our analyses by using different functions per lottery-problem domain.

[3] It is important to not confuse gain-loss asymmetries and the loss-aversion parameter λ included in CPT: The former consists of observable behaviors whereas the latter is one possible account of said behavior. As we will see later on, other theories characterize this behavior differently.

[4] As discussed in detail by Ingersoll (2008), the probability-weighting function in proposed by Kahneman and Tversky (1979; see Equation 6) is not monotonic for values below .278.

Table 1
Examples of choice paradoxes, the qualitative predictions associated with CPT.

Choice Paradox			Qualitative Predictions		Out-of-sample Predictions	
			CPT	P($\mathcal{A}$)	TAX	DFT$_e$
Event Splitting/Coalescing (ESC)	$\mathcal{A} = \begin{pmatrix}\$100 & \$50\\ .85 & .15\end{pmatrix}$	$\mathcal{B} = \begin{pmatrix}\$100 & \$7\\ .95 & .05\end{pmatrix}$	$\mathcal{A} \Leftrightarrow \mathcal{A}'$	.74	.50	.62
	$\mathcal{A}' = \begin{pmatrix}\$100 & \$50 & \$50\\ .85 & .10 & .05\end{pmatrix}$	$\mathcal{B}' = \begin{pmatrix}\$100 & \$100 & \$7\\ .85 & .10 & .05\end{pmatrix}$		.38	.44	.49
Stochastic Dominance (SD)	$\mathcal{A}'' = \begin{pmatrix}\$96 & \$14 & \$12\\ .90 & .05 & .05\end{pmatrix}$	$\mathcal{B}'' = \begin{pmatrix}\$96 & \$90 & \$12\\ .85 & .05 & .10\end{pmatrix}$	$\mathcal{A}''$	.27	.46	.37
Upper Tail Independence (UTI)	$\mathcal{A} = \begin{pmatrix}\$110 & \$44 & \$40\\ .80 & .10 & .10\end{pmatrix}$	$\mathcal{B} = \begin{pmatrix}\$110 & \$96 & \$10\\ .80 & .10 & .10\end{pmatrix}$	$\mathcal{A} \Leftrightarrow \mathcal{A}''$	.33	.45	.47
	$\mathcal{A}'' = \begin{pmatrix}\$96 & \$44 & \$40\\ .80 & .10 & .10\end{pmatrix}$	$\mathcal{B}'' = \begin{pmatrix}\$96 & \$10\\ .90 & .10\end{pmatrix}$		.67	.51	.56
Lower Cumulative Independence (LCI)	$\mathcal{A} = \begin{pmatrix}\$52 & \$48 & \$3\\ .10 & .10 & .80\end{pmatrix}$	$\mathcal{B} = \begin{pmatrix}\$96 & \$12 & \$3\\ .05 & .05 & .90\end{pmatrix}$	$\mathcal{A} \Rightarrow \mathcal{A}''$	.62	.52	.51
	$\mathcal{A}'' = \begin{pmatrix}\$52 & \$12\\ .10 & .90\end{pmatrix}$	$\mathcal{B}'' = \begin{pmatrix}\$96 & \$12\\ .05 & .95\end{pmatrix}$		.26	.45	.38
Upper Cumulative Independence (UCI)	$\mathcal{A} = \begin{pmatrix}\$110 & \$44 & \$40\\ .80 & .10 & .10\end{pmatrix}$	$\mathcal{B} = \begin{pmatrix}\$110 & \$98 & \$10\\ .80 & .10 & .10\end{pmatrix}$	$\mathcal{B} \Rightarrow \mathcal{B}''$	.30	.44	.47
	$\mathcal{A}'' = \begin{pmatrix}\$98 & \$40\\ .80 & .20\end{pmatrix}$	$\mathcal{B}'' = \begin{pmatrix}\$98 & \$10\\ .90 & .10\end{pmatrix}$		.58	.51	.59

Note. Column ‘ Qualitative Predictions CPT’ indicates the preference constraints predicted by CPT irrespective of the parametric form of its subjective representations. The values in column ‘ P($\mathcal{A}$)’ indicate the proportion of $\mathcal{A}/\mathcal{A}'/\mathcal{A}''$ choices made for lottery problem in ‘Choice Paradox’, with values taken from Birnbaum (2005), Birnbaum (2008), Birnbaum and Navarrete (1998). Columns referring to TAX and DFT$_e$ report the average ‘out-of-sample’ quantitative predictions for TAX and DFT$_e$ based on the individual parameters estimates obtained from the data from Rieskamp (2008, Study 2). Note that the lottery problems listed here were not included in the Rieskamp’s (2008) study, hence the predictions being ‘out-of-sample’.

assumptions that underlie the target model, in the present case CPT. Please note that preferences, whether paradoxical or not, consist of ordinal relations between alternatives. In order to test the presence of any set of preferences based on observed choices, it is necessary to take into account the stochastic nature of the latter (e.g., Luce, 1953; Rieskamp, 2008). Under minimalist assumptions, the preference $\mathcal{A} \succ \mathcal{B}$ results in the qualitative prediction that the probability of $\mathcal{A}$ being chosen over $\mathcal{B}$, $P(\mathcal{A} \succ \mathcal{B})$, lies between $\frac{1}{2}$ and 1. This means that when preference $\mathcal{A} \succ \mathcal{B}$ is assumed, scenarios in which most participants in a large sample choose $\mathcal{B}$ over $\mathcal{A}$ are deemed extremely unlikely, suggesting a failure of the model. For example, the probability of $\mathcal{B}$ being chosen over $\mathcal{A}$ at least 37 times out of fifty choices (74%) is at best .0005, when assuming that $P(\mathcal{A} \succ \mathcal{B}) = \frac{1}{2}$.

We will now review some of the paradoxes reported by Birnbaum (2008). Their names refer to the property required under CPT that is violated. The paradoxes selected are sometimes referred to as ‘*strong paradoxes*’ as they do not hinge on the components of CPT taking any specific parametric and/or functional forms, such as an inverse-S-shaped probability-weighting function like the ones illustrated in Fig. 1 (e.g., Birnbaum, 2005). Empirical demonstrations of the different choice paradoxes discussed here are provided in Table 1. When describing the paradoxes, we will rely on pure-gain lotteries (i.e., all events yield non-negative outcomes), with $z^\uparrow > y^\uparrow > y > x^\uparrow > x > x^\downarrow > z > 0$, although the properties discussed here also hold in the domains of pure-loss and mixed lotteries. Outcomes with superscripts ↑ and ↓ are used here in order to make the relative rank of outcomes as clear as possible in cases where outcomes are improved or worsened (e.g., $x^\downarrow$ is worse than x and $x^\uparrow$ is better than x). Also, common outcomes across lotteries are presented in bold. Finally, let p, q, r, and s denote probabilities associated with the events of lotteries.

2.1. Event splitting and coalescing

We begin with the CPT assumption of event coalescing, according to which the evaluation of lotteries should be invariant to the form that they are presented, specifically the way outcomes are associated with events. For example, lottery $\mathcal{A} = \begin{pmatrix}\$100 & \$50\\ .85 & .15\end{pmatrix}$ should be perceived as equivalent to $\mathcal{A}' = \begin{pmatrix}\$100 & \$50 & \$50\\ .85 & .10 & .05\end{pmatrix}$ or $\mathcal{A}'' = \begin{pmatrix}\$100 & \$100 & \$50\\ .45 & .40 & .15\end{pmatrix}$. Several studies have shown this invariance assumption to fail, with lotteries splitting the events with the worst outcomes, such as $\mathcal{A}'$, being seen as less attractive than $\mathcal{A}$. Also, these studies show lotteries splitting the events with best outcomes like $\mathcal{A}''$ being perceived as more attractive (Starmer & Sugden, 1993; Birnbaum & Navarrete, 1998; Humphrey, 1995, Humphrey, 2001). The specific example given in Table 1 also speaks against this form invariance, as most individuals prefer $\mathcal{A}$ over $\mathcal{B}$ at the same time that most also prefer $\mathcal{B}'$ over $\mathcal{A}'$.

2.2. First-order stochastic dominance

Event-splitting and coalescing operations are part of a recipe that produces (non-transparent) tests of first-order stochastic dominance (Birnbaum, 1997). A lottery $\mathcal{A}$ stochastically dominates a lottery $\mathcal{B}$ when the probability of receiving a value of at least x in the former is always larger or equal than in the latter. First, let us consider lottery $\mathcal{A} = \begin{pmatrix} y & x \\ p & 1-p \end{pmatrix}$. Now, let us split the event associated with one outcome, for example y, in order to create a new lottery $\mathcal{A}' = \begin{pmatrix} y & x & x \\ p-s & s & 1-p \end{pmatrix}$. According to event coalescing, $\mathcal{A} \sim \mathcal{A}'$. Now, let us make one of the x outcomes better, yielding $\mathcal{A}'' = \begin{pmatrix} y & x^{\uparrow} & x \\ p-s & s & 1-p \end{pmatrix}$. Moreover, let us also create another lottery based on $\mathcal{A}$, by splitting the event with the best outcome x into two and making one of the outcomes worse, resulting in $\mathcal{B}'' = \begin{pmatrix} y & y^{\downarrow} & x \\ p-q & q & 1-p \end{pmatrix}$. Monotonicity implies that $\mathcal{A}'' \prec \mathcal{A}$ and $\mathcal{A} \prec \mathcal{B}''$. In turn, transitivity implies that $\mathcal{A}'' \prec \mathcal{B}''$. [5] Note that $\mathcal{A}''$ stochastically dominates $\mathcal{B}''$. The empirical results shown in Table 1 indicate that most individuals turn out to choose the dominated option $\mathcal{B}''$ over $\mathcal{A}''$.

2.3. Upper tail independence

This property, originally described by Green and Jullien (1988) and subsequently tested by Wu (1994), is a necessary condition for rank-dependent utility models such as CPT to hold. Specifically, changes in the (common) highest-ranking outcomes do not affect preferences as long as ranks are preserved:

$$\mathcal{A} = \begin{pmatrix} \mathbf{z}^{\uparrow} & x & x^{\downarrow} \\ p & q & 1-p-q \end{pmatrix} \quad \mathcal{B} = \begin{pmatrix} \mathbf{z}^{\uparrow} & y & z \\ p & r & 1-p-r \end{pmatrix}.$$

$$\mathcal{A}' = \begin{pmatrix} \mathbf{y}^{\uparrow} & x & x^{\downarrow} \\ p & q & 1-p-q \end{pmatrix} \quad \mathcal{B}' = \begin{pmatrix} \mathbf{y}^{\uparrow} & y & z \\ p & r & 1-p-r \end{pmatrix}.$$

2.4. Cumulative independence

Lower and upper cumulative independence are properties that need to be satisfied by any model, like CPT, that jointly requires *comonotonic independence*, *monotonicity*, *transitivity*, and *coalescing* to hold (Birnbaum, 1997). Both independence properties should be observable by manipulating the lotteries' worst (lower) and best (upper) outcomes. According to CPT, a specific implication at the level of preferences follows from these manipulations.

2.4.1. Lower cumulative independence

$$\mathcal{A} = \begin{pmatrix} y & x & \mathbf{z} \\ p & q & r \end{pmatrix} \quad \mathcal{B} = \begin{pmatrix} y^{\uparrow} & x^{\downarrow} & \mathbf{z} \\ p & q & r \end{pmatrix},$$

$$\mathcal{A}'' = \begin{pmatrix} y & \mathbf{x}^{\downarrow} \\ p+q & r \end{pmatrix} \quad \mathcal{B}'' = \begin{pmatrix} y^{\uparrow} & \mathbf{x}^{\downarrow} \\ p & q+r \end{pmatrix}.$$

According to comonotonic independence, replacing the common outcome z in $\mathcal{A}$ and $\mathcal{B}$ with outcome $x^{\downarrow}$ with the same rank (they correspond to the worst outcome in both lotteries) preserves the preference. Now, let us improve $\mathcal{A}$ by changing x with y. These two changes result in the improved lotteries $\mathcal{A}' = \begin{pmatrix} y & y & \mathbf{x}^{\downarrow} \\ p & q & r \end{pmatrix}$ and $\mathcal{B}' = \begin{pmatrix} y^{\uparrow} & \mathbf{x}^{\downarrow} & \mathbf{x}^{\downarrow} \\ q & r & p \end{pmatrix}$. Finally, apply coalescing to both $\mathcal{A}'$ and $\mathcal{B}'$, obtaining $\mathcal{A}''$ and $\mathcal{B}''$, respectively. According to CPT, a preference for $\mathcal{A}$ over $\mathcal{B}$ implies a preference for $\mathcal{A}''$ over $\mathcal{B}''$. The results reported in Table 1 show that the proportion of people choosing $\mathcal{A}''$ is smaller than $\mathcal{A}$, a pattern that is at odds with the postulated implication.

2.4.2. Upper cumulative independence

$$\mathcal{A} = \begin{pmatrix} \mathbf{z}^{\uparrow} & y & x \\ p & q & r \end{pmatrix} \quad \mathcal{B} = \begin{pmatrix} \mathbf{z}^{\uparrow} & y^{\uparrow} & x^{\downarrow} \\ p & q & r \end{pmatrix},$$

$$\mathcal{A}'' = \begin{pmatrix} \mathbf{y}^{\uparrow} & x \\ p+q & r \end{pmatrix} \quad \mathcal{B}'' = \begin{pmatrix} \mathbf{y}^{\uparrow} & x^{\downarrow} \\ p & q+r \end{pmatrix}.$$

[5] Transitivity holds that if for any three options $\mathcal{A}$, $\mathcal{B}$, and C, if $\mathcal{A} \succ \mathcal{B}$, and $\mathcal{B} \succ C$, then $\mathcal{A} \succ C$. Empirical evidence supporting transitivity can be found in Regenwetter, Dana, and Davis-Stober (2011), and Birnbaum and Schmidt (2010).

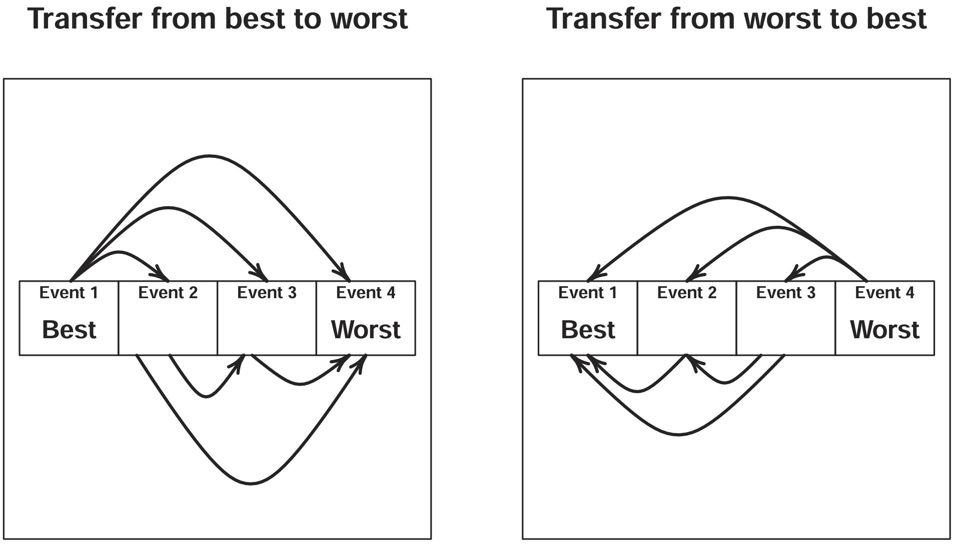

Fig. 2. Illustration of TAX's weight transfer schemes across ranked-ordered events.

Again, according to comonotonic independence, we can exchange the common outcome $z^{\uparrow}$ for another outcome of the same rank, in this case reducing it to $y^{\uparrow}$ (they correspond to the best outcomes in both lotteries). Also, let us decrease y to x in $\mathcal{A}$. These two changes yield $\mathcal{A}' = \begin{pmatrix} \mathbf{y}^{\uparrow} & x & x \\ p & q & r \end{pmatrix}$and $\mathcal{B}' = \begin{pmatrix} \mathbf{y}^{\uparrow} & \mathbf{y}^{\uparrow} & x^{\downarrow} \\ p & q & r \end{pmatrix}$. Coalescing of $\mathcal{A}'$ and $\mathcal{B}'$ yields $\mathcal{A}''$ and $\mathcal{B}''$, respectively. According to CPT, the preference for $\mathcal{B}$ implies a preference for $\mathcal{B}''$. Once again, the data reported in Table 1 are at odds with this prediction as less people chose $\mathcal{B}''$ over $\mathcal{A}''$ than $\mathcal{B}$ over $\mathcal{A}$.

3. Transfer of attention exchange model

CPT's failure to account for Birnbaum's 'new choice paradoxes' indicates the need to consider an alternative theory that can account for the violation of these properties under certain circumstances. Birnbaum and colleagues proposed the TAX model, which assumes a weighting of events in which attention is transferred between events. TAX is a member of a class of *configural-weighting models* whose development goes beyond the study of choices under risk and long precedes the development of CPT (see Birnbaum, 1973; Birnbaum, Parducci, & Gifford, 1971; Birnbaum & Stegner, 1979; for a review, see Birnbaum, 2008). According to TAX, weights are transferred between the different branches based on the ranks of their associated outcomes (for an illustration, see Fig. 2). For instance, for $\mathcal{A} = \begin{pmatrix} \$100 & \$50 & \$10 \\ .60 & .20 & .20 \end{pmatrix}$, attention weighting can be transferred from the higher-ranked outcomes to the lower-ranked ones (e.g., from \$100 to both \$50 and \$10, and from \$50 to \$10). The utility of a lottery $\mathcal{L}$ is based on the subjective representations of outcomes and probabilities, $u(x)$ and $\rho(p)$, respectively, and an *attention-exchange function*, denoted $\omega(p_k, p_j, K)$. This function determines how much weight gets transferred between an event associated with a higher-ranked outcome x_k and an event associated with a lower-ranked x_j outcome. The overall value of a lottery is then determined by this transfer (scaled by the weighted probabilities):

$$V(\mathcal{L}) = \frac{\sum_{k=1}^{K} u(x_k)\rho(p_k) + \sum_{k=1}^{K} \sum_{j=1}^{k} (u(x_k) - u(x_j))\omega(p_k, p_j, K)}{\sum_{k=1}^{K} \rho(p_k)}. \tag{7}$$

In contrast with CPT, which ranks gains and losses separately, TAX assumes that all events are jointly ranked in mixed lotteries. In the case of pure-gain and mixed lotteries, they are ranked from best to worst. In the case of pure-loss lotteries, events are ranked from worst to best (see Birnbaum & Bahra, 2007). Luce and Marley (2005) and Marley and Luce (2005) provide a thorough characterization of the general class of TAX models and discuss some open problems (see also Birnbaum, 1997).

Also, TAX does not assume event coalescing, such that event-splitting and coalescing operations lead to changes in the value of the lotteries. For example, let us split the event with the best outcome in $\mathcal{A}$, resulting in $\mathcal{A}' = \begin{pmatrix} \$100 & \$100 & \$50 & \$10 \\ .30 & .30 & .20 & .20 \end{pmatrix}$. If there is a transfer from better to worse events, the utility of $\mathcal{A}'$ is greater than $\mathcal{A}$ because some of the weighting is transferred between the split events. If we instead split the event with the worst outcome, resulting in $\mathcal{A}'' = \begin{pmatrix} \$100 & \$50 & \$10 & \$10 \\ .60 & .20 & .10 & .10 \end{pmatrix}$, a larger part of the weighting will be attributed to these two events associated with the \$10 outcome. In fact, TAX attributes behavioral phenomena such as the Allais' Common Consequence Paradox to operations such as the splitting and coalescing of events (for a detailed discussion, see Birnbaum, 2004).

According to the parametric version of TAX that is typically adopted (Birnbaum, 2008; Scheibehenne & Pachur, 2015), the subjective representation of outcomes is equivalent to CPT, but without a loss-aversion parameter ($\lambda = 1$).[6] The observed differences between gains and losses usually associated with the loss-aversion parameter λ are instead attributed to the particular way attention is transferred between events. Loss aversion is often operationalized as a dislike for mixed symmetrical lotteries such as $\mathcal{A}' = \begin{pmatrix} \$100 & -\$100 \\ .50 & .50 \end{pmatrix}$. But according to TAX, 'losses loom larger than gains' because attention is being transferred from the best outcome (\$100) to the worst (− \$100), the same kind of transfer that would occur between the best and worst outcomes of lottery $\begin{pmatrix} \$100 & \$20 \\ .50 & .50 \end{pmatrix}$ and would lead individuals to dislike it relative to the option of receiving \$60 for sure.[7]

The subjective representation of probabilities is captured by a power function ρ with parameter γ, with $\rho(p) \geqslant p$ when $0 \leqslant \gamma \leqslant 1$ and $\rho(p) < p$ when $\gamma > 1$:

$$\rho(p) = p^{\gamma}. \tag{8}$$

In the attention-transfer function ω, the amount of weight transferred between any two branches is a fixed proportion of the subjective probability of the branch losing weight:

$$\omega(p_i, p_k, K) = \begin{cases} \frac{\delta\rho(p_i)}{K+1}, & \text{for } \delta > 0, \\ \frac{\delta\rho(p_k)}{K+1}, & \text{for } \delta \leqslant 0. \end{cases} \tag{9}$$

Parameter δ determines the direction and magnitude of the transfer: When $\delta > 0$, weighting is transferred from branches with higher-ranked events to branches with the lower-ranked ones. When $\delta < 0$, weighting is transferred in the opposite direction instead.[8] Both cases are illustrated in Fig. 2. Finally, as in the case of CPT, a logistic choice rule is used in TAX to capture the probabilistic aspect of individuals' choices (i.e., Eq. (5) applies to TAX as well). Using the 'prior' parameter values $\alpha = 1$, $\gamma = 0.7$, and $\delta = 1$, Birnbaum and colleagues showed that TAX was able to predict the modal responses associated with the choice paradoxes discussed above (for a review, see Birnbaum, 2008).

4. Decision field theory

We now turn to another model that could, in principle, account for Birnbaum's choice paradoxes. *Decision Field Theory* (DFT; Busemeyer & Townsend, 1993) is a sequential-sampling model that attempts to characterize the dynamic nature of choices and the several effects that follow from it (see Busemeyer & Diederich, 2002; Busemeyer & Townsend, 1993; Diederich, 2003a, Diederich, 2003b; Dror, Basola, & Busemeyer, 1999; Hotaling & Busemeyer, 2012; Johnson & Busemeyer, 2005; Scheibehenne, Rieskamp, & González-Vallejo, 2009; see also Birnbaum & Jou, 1990).

In a nutshell, DFT assumes that the individual decision maker deliberates by thinking about the possible events in each lottery and their associated outcomes. This deliberative process is comprised of a sequence of samples from the different options by means of mental simulation (for similar views, see Griffiths, Vul, & Sanborn, 2012; Stewart & Simpson, 2008; see also the "mental simulation heuristic" discussed by Kahneman and Tversky, 1982). At every time point t an event from each lottery $\mathcal{A}$ and $\mathcal{B}$ are sampled; e.g., $\hat{x}_{\mathcal{A}}$ and $\hat{x}_{\mathcal{B}}$. Events from each lottery are sampled proportionally to its respective probability p. The difference between $\hat{x}_{\mathcal{A}}$ and $\hat{x}_{\mathcal{B}}$ is then computed and added to a running tally:

$$\Delta V_t = u(\hat{x}_{\mathcal{A}}) - u(\hat{x}_{\mathcal{B}}) + \xi_t + \Delta V_{t-1}, \tag{10}$$

where ξ_t is a noise term that follows a Normal distribution with mean 0 and variance σ^2. Note that the expected ΔV corresponds to the difference between the two lotteries' utilities under Expected Utility Theory (see Eq. (1)). The sampling process is assumed to take place until ΔV reaches one of two response thresholds, producing a response. Fig. 3 illustrates this sampling process for lotteries $\mathcal{A} = \begin{pmatrix} \$350 & \$40 \\ .10 & .90 \end{pmatrix}$ and $\mathcal{B} = \begin{pmatrix} \$100 & \$20 \\ .50 & .50 \end{pmatrix}$. Samples being taken until the ΔV tally crosses the upper boundary, leading to option $\mathcal{A}$ being chosen (if the lower boundary had been crossed instead, option $\mathcal{B}$ would have been chosen).

[6] For convenience, we will refer to this special parametric version of TAX as simply 'TAX', unless a distinction between the parametric version and the general case is necessary.

[7] It should be noted that the different ways TAX is assumed to transfer weights depending on the type of lottery, especially when assuming the same parameter values across lottery types, can lead to paradoxical predictions: If pure-loss lotteries are ranked from worst to best, and mixed lotteries from best to worst (Birnbaum & Bahra, 2007), with $\alpha = 1$, $\gamma = 0.70$, $\delta = 1$ then it follows that lotteries $\begin{pmatrix} \$5 & -\$12 & -\$14 & -\$96 \\ .05 & .05 & .05 & .85 \end{pmatrix}$ and $\begin{pmatrix} \$0 & -\$12 & -\$14 & -\$96 \\ .05 & .05 & .05 & .85 \end{pmatrix}$ have subjective values -75.90 and -32.80. In short, improving one outcome such that pure-loss lottery becomes a mixed lottery makes the lottery worse. A similar scenario can be found in lottery problems 13 and 19 (lottery $\mathcal{F}$) reported in Table 3 of Birnbaum and Bahra (2007). One way to try to avoid such problems, while keeping the rank orderings proposed by Birnbaum and Bahra, is to allow for different parameter values across lottery types.

[8] One important notion in TAX is that no weight is created or destroyed during the transfer process. With the weight-transfer function described in Eq. (9), this is only achieved if $-\frac{K_{max}+1}{K_{max}-1} \leqslant \delta \leqslant \frac{K_{max}+1}{K_{max}-1}$, where K_{max} is the maximum number of events a lottery can have. In the novel experiment reported here, lotteries have up to five events, which means that $-1.5 \leqslant \delta \leqslant 1.5$. We will constraint δ to take only take on values from this range. For consistency, this constraint will also be applied when fitting TAX to the data from Rieskamp (2008).

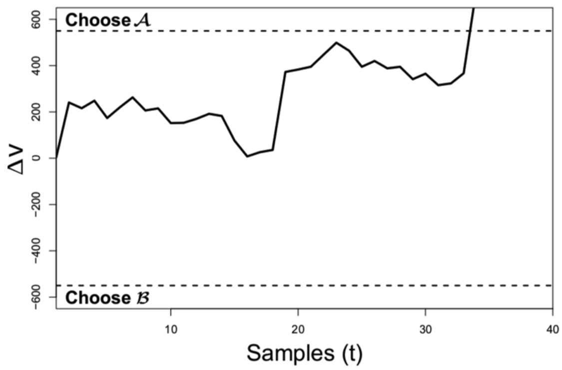

Fig. 3. Illustration of DFT's sequential sampling process. The dashed lines correspond to the response thresholds.

When applied to risky choice data alone (i.e., no response-time data being considered), a *simplified version* of DFT is often used (Busemeyer & Townsend, 1993, 439–440; see also Bhatia, 2014; Rieskamp, 2008). According to this model, the *expected value* of each lottery corresponds to the value that would obtained with Expected Utility Theory:[9]

$$\mathbb{E}(V(\mathcal{L})) = \sum_{k=1}^{K} u(x_k)\cdot\mathbb{E}(w_k). \tag{11}$$

where w_k are the sampling weights associated with each event, with $\mathbb{E}(w_k) = p_k$. This simplified DFT model includes a choice rule that unlike Eq. (6), not only takes into consideration each lotteries' valuation, but also their variances, with $\sigma_{\mathcal{L}}^2 = \sum_{i=1}^{K} p_i(u(x_i) - \mathbb{E}(V(\mathcal{L})))^2$. Specifically, responses become more deterministic as lottery variance decreases:[10]

$$P(\mathcal{A} \succ \mathcal{B}) = \frac{1}{1 + \exp\left(-2\theta\frac{\mathbb{E}(V(\mathcal{A})) - \mathbb{E}(V(\mathcal{B}))}{\sqrt{\sigma_{\mathcal{A}}^2 + \sigma_{\mathcal{B}}^2}}\right)}, \tag{12}$$

with parameter θ capturing the sensitivity of choice probabilities to differences between the lotteries' subjective values. Despite its many merits, DFT in its original form is unable to account for the many of the choice phenomena successfully captured by Prospect Theory, CPT and/or TAX (Bhatia, 2014; Rieskamp, 2008; see also Birnbaum & Jou, 1990). In fact, it makes the same qualitative predictions as EUT. The reason is that constructs like probability weighting or sign-rank-dependent decision weighting are simply not amenable to the sampling process postulated by the model in its original form.

As an example, let us see how CPT, TAX, and EUT/DFT are able to accommodate individuals valuations of lotteries, and how these vary as a function of event probabilities. The data considered here come from a study reported by Gonzalez and Wu (1999), in which participants were asked to judge the monetary value (i.e., their certainty equivalents) of binary lotteries. We fitted each model to the data by minimizing SSE. Fig. 4 plots the average values of lotteries $\mathcal{A} = \begin{pmatrix} y & 0 \\ p & 1-p \end{pmatrix}$ and $\mathcal{B} = \begin{pmatrix} y & x \\ p & 1-p \end{pmatrix}$ as probability p varies between.01 and.99. It is clear that EUT/DFT grossly fails to account for the data, whereas CPT and TAX are able to provide a much closer fit. These performance differences are due to the fact that the weighting of events in EUT/DFT, unlike CPT/TAX, cannot deviate from their respective probabilities.

To overcome these limitations, Bhatia (2014) proposed an extension of DFT, which we will refer to as DFT$_e$, that introduces the possibility of *non-proportional sampling*. At every time point t, with probability π the individual succeeds in sampling the events proportionally. But with probability $1 - \pi$, the events are sampled with equal probability (i.e., $\frac{1}{K}$). This non-proportional sampling can be interpreted as a focus on the potential events while disregarding their respective probabilities. Instead of having the same expectations as Expected Utility Theory, DFT$_e$ assumes that the valuation of a lottery $\mathcal{L}$ consists of a mixture of two sampling regimes with weights π and $1 - \pi$: [11]

$$\mathbb{E}(V(\mathcal{L})) = \sum_{k=1}^{K} \pi\cdot u(x_k)\cdot\mathbb{E}(w_k) + \sum_{k=1}^{K} (1-\pi)\cdot u(x_k)\cdot\frac{1}{K}. \tag{13}$$

[9] The present description of DFT does not discuss the final DFT model discussed by Busemeyer and Townsend (1993), which includes features such as anchor points, decaying growth rates, among others (for reference, our description corresponds to their 'Stage 3'; see their Table 2). Our focus on a simplified version of DFT is not problematic, as the full model will generally predict a preference for the option with the highest expected utility (exceptions can occur when response thresholds are extremely small; see Bhatia, 2014, p. 1098).

[10] In the present case, we are assuming that the events of each lottery are *independent*. If they covary, then $\sigma^2 = \sigma_{\mathcal{A}}^2 + \sigma_{\mathcal{B}}^2 - 2\text{cov}_{\mathcal{A},\mathcal{B}}$. The introduction of event covariances in DFT leads to empirical predictions that have been supported empirically (Andraszewicz, Rieskamp, & Scheibehenne, 2015).

[11] For somewhat similar proposals, see Viscusi (1989) and Mukherjee (2010).

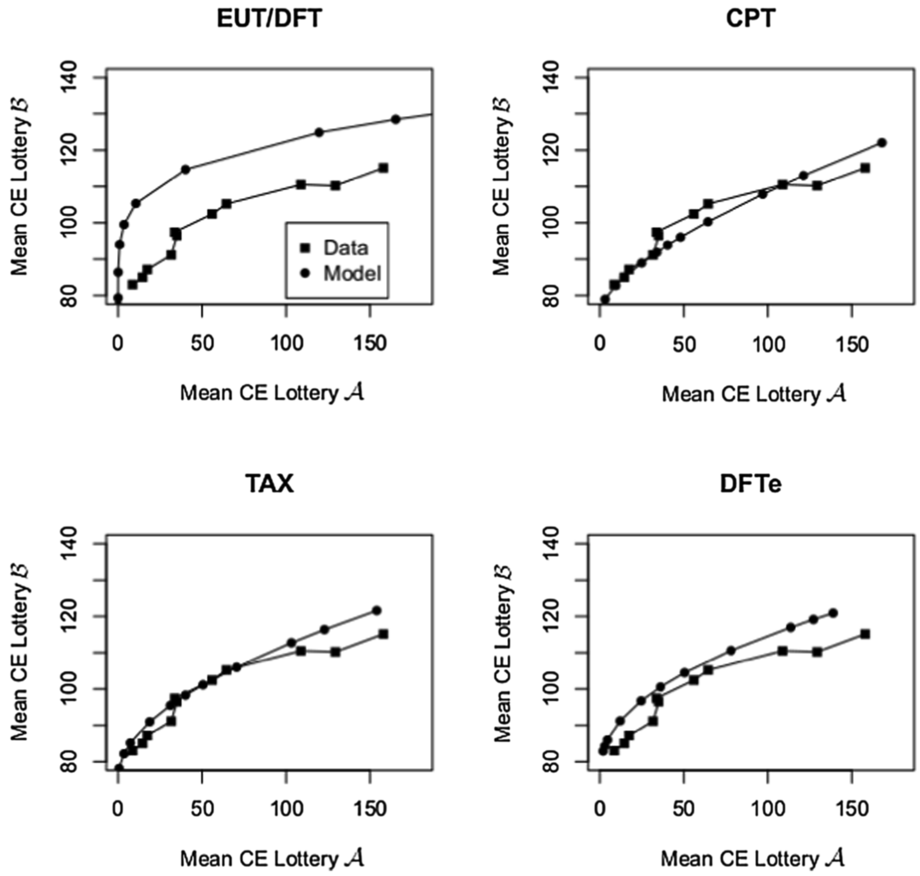

Fig. 4. Gonzalez and Wu's (1999), mean certainty-equivalent (CE) judgments for lotteries $\mathcal{A} = \begin{pmatrix} y & \$0 \\ p & 1-p \end{pmatrix}$ and $\mathcal{B} = \begin{pmatrix} y & x \\ p & 1-p \end{pmatrix}$ (averaged across x and y). Each point corresponds to a given probability p, which took on values.01,.05,.10,.25,.40,.50,.60,.75,.90,.95, and.99.

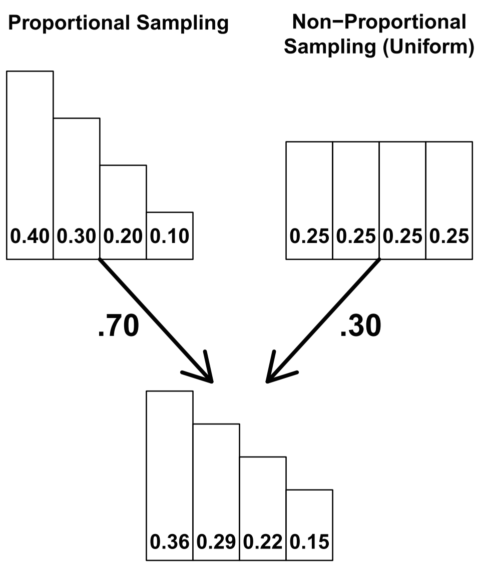

Fig. 5. Illustration of DFT_e's weighting via mixture sampling with $\pi = .70$ in a four-event lottery with probabilities .40, .30, .20, and .10.

As shown in Fig. 5, the occurrence of non-proportional (in this case equiprobable) sampling effectively leads to an overweighting of small probabilities and an underweighting of large probabilities, similar to CPT. Rare events therefore get sampled more often than expected, and highly-likely events less often. Keep in mind that in DFT$_e$, outcome rank is completely irrelevant, and events are only over/underweighted to the extent that their respective probabilities are above/below $\frac{1}{K}$. As shown in Fig. 4, the possibility of non-proportional sampling improves DFT's ability to characterize the Gonzalez and Wu (1999) data considerably.

Using individual choice data from Rieskamp (2008), Bhatia (2014) showed that DFT$_e$ systematically outperforms the simplified DFT model it extends. Bhatia also showed that DFT$_e$ can account for the modal responses of the lottery problems originally reported by Kahneman and Tversky (1979), including the Allais' paradoxes: in the case of Allais' Common Consequence Paradox, the non-proportional sampling will overweight the small-probability $0 event in $\mathcal{B}$ while having no effect on $\mathcal{A}$, leading to a preference for the latter. In contrast, the potential gains in both $\mathcal{A}'$ and $\mathcal{B}'$ have probabilities below $\frac{1}{2}$, which means that both benefit from non-proportional sampling, which in turn leads to a preference for $\mathcal{B}'$.

At this point, there is no strong theoretical commitment to the causes of the non-proportional sampling, as one can think of different (non-mutually exclusive) reasons. One possibility is that it is due to a momentary failure to keep in mind the different probabilities. Another possibility is that it is produced by *errors* in the identification of the sampled event. This interpretation of the postulated non-proportional sampling is particularly interesting given that sampling errors have been used to explain a wide range of phenomena (e.g., Costello & Watts, 2014; Dougherty, Gettys, & Ogden, 1999; Erev, Wallsten, & Budescu, 1994; Lin, Donkin, & Newell, 2015). For example, in the domain of probability judgments, recent work by Hilbert (2012) and later Costello and Watts (2014), Costello and Watts (2016) has shown that many well-known phenomena such as conjunction fallacies (e.g., the 'Linda effect') can be understood as the byproduct of a faulty sampling process.

One limitation of Bhatia's DFT$_e$ implementation is that it is unable to accommodate gain-loss asymmetries, which is traditionally achieved by CPT via its loss aversion parameter λ and by TAX via the transfer of attention across ranked outcomes. Fortunately, the sampling process postulated by DFT$_e$ can be easily extended in order to enable the characterization of potential asymmetries between gains and losses: Let τ be an attention-weight parameter ranging from -1 to 1 that biases sampling whenever proportional sampling fails. Instead of all events then being sampled with equal probability, events associated with gains and loss outcomes might be sampled at different rates. The idea here is that when individuals focus on events while disregarding their respective probabilities, a greater deal of attention is given to either the gains or the losses. Indeed, there is a considerable body of work showing that negative information is more attention grabbing (e.g., Hamilton & Zanna, 1972; Pratto and John, 1991). For a mixed lottery with M and N events associated with loss and gain (or zero) outcomes with probabilities p_m^- and p_n^+, their respective sampling probabilities are

$$\pi p_m^- + (1-\pi)\frac{1+\tau}{M(1+\tau)+N(1-\tau)}, \quad \text{and}$$
$$\pi p_n^+ + (1-\pi)\frac{1-\tau}{M(1+\tau)+N(1-\tau)}.$$

For example, when $\pi = .70$ and $\tau = 0.75$, the events of mixed lottery $\mathcal{L} = \begin{pmatrix} \$100 & \$50 & -\$50 & -\$100 \\ .25 & .25 & .25 & .25 \end{pmatrix}$ are sampled with probabilities .31, .31., .19, and .19, with gains being overweighted relative to losses. The opposite weighting of .19, .19, .31, and .31 is obtained with $\tau = -0.75$. Regarding choice paradoxes, Bhatia (2014) argued that event-splitting and coalescing effects could, in principle, be accounted for by DFT$_e$'s mixed sampling process (for an earlier discussion, see Birnbaum, 1999): By splitting events, one is effectively increasing the probability that the outcome associated with the split event is sampled in the case of non-proportional sampling. This increase leads to changes in the overall evaluation of the lottery: If the event associated with the best outcome is split, $V(\mathcal{L})$ increases. Alternatively, if the event associated with the worst outcome is split instead, $V(\mathcal{L})$ will decrease. For example, with $\alpha = \pi = 0.7$, the subjective values of lotteries $\mathcal{A} = \begin{pmatrix} \$100 & \$50 \\ .85 & .15 \end{pmatrix}$, $\mathcal{A}' = \begin{pmatrix} \$100 & \$50 & \$50 \\ .85 & .10 & .05 \end{pmatrix}$, and $\mathcal{A}'' = \begin{pmatrix} \$100 & \$100 & \$50 \\ .45 & .40 & .15 \end{pmatrix}$ are, in order, 22.66, 22.17, and 23.14. No further 'new' paradoxes were considered by Bhatia (2014).

5. Current state of affairs

Despite the strong empirical case against CPT, a paradigm shift towards an alternative model like TAX *has not* taken place, with CPT's popularity remaining largely unaffected. This state of affairs can be partially attributed to the fact that most researchers fit CPT to choice data concerning binary lotteries, an experimental-design option that leaves out the multiple-outcome lotteries necessary to demonstrate the occurrence of choice paradoxes like Birnbaum's. One problem with this situation is that the different models provide rather different characterizations of choice behavior: For instance, CPT describes gain-loss asymmetries via the notion of loss aversion, whereas TAX attributes these asymmetries to an attention-exchange process. By continuing to rely on a model that has been consistently rejected, one is very likely to be led astray (for a relevant discussion, see Rotello, Heit, & Dubé, 2015).

On the other hand, the studies demonstrating the occurrence of choice paradoxes typically involve a very limited set of tailored lottery problems. This aspect should not be overlooked given the existence of reports suggesting that experimental-design options of this kind can affect participants' choices and its characterization (for a drastic example, see Spektor, Kellen, and Hotaling, 2018). For instance, Millroth, Nilsson, and Juslin (2018) showed that CPT's value and probability-weighting functions estimated from individuals encountering a single lottery problem versus multiple lottery problems differ substantially. Also, Stewart, Reimers, and Harris (2014) showed that the shape of the value and probability-weighting functions can be drastically affected by the characteristics of the lottery problems (see also Walasek & Stewart, 2015).

These reports indicate the need to assess the joint occurrence of the different choice paradoxes in a larger and more diverse set of

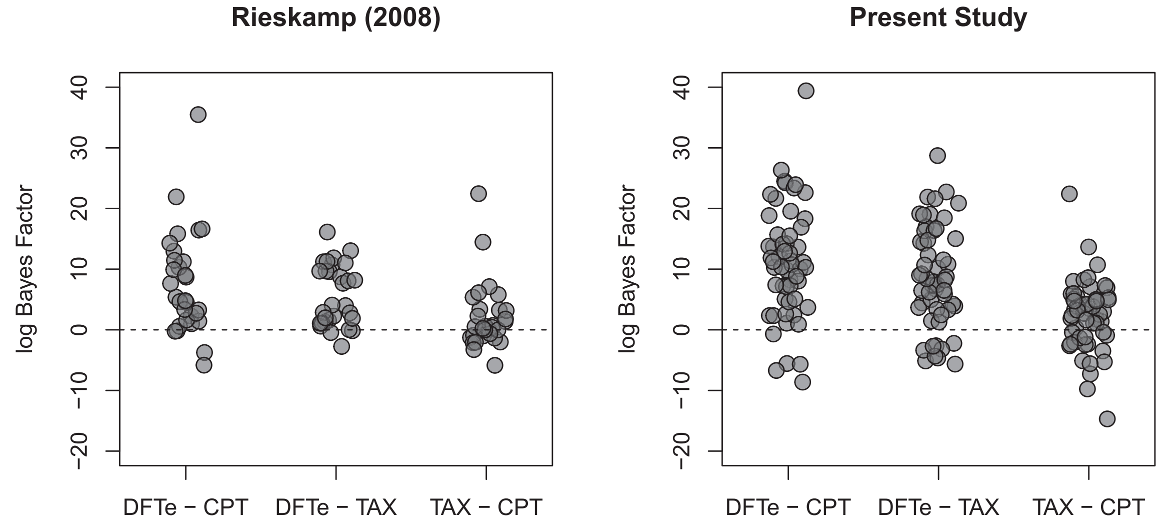

Fig. 6. Log Bayes Factors for different model pairings. Each point corresponds to one individual participant. Left Panel: Rieskamp (2008). Right Panel: Present Study.

lottery problems (for a similar point, see Birnbaum, 2008, p. 497). A recent study by Birnbaum and Schmidt (2015) showed a reduction in the occurrence of choice paradoxes among more experienced participants, a result that reinforces the need for further testing. Also relevant is a study by Erev, Ert, Plonsky, Cohen, and Cohen (2017) attempting to jointly replicate a large number of established behavioral phenomena where notably, no event-splitting/coalescing effects were observed. Moreover, it is not clear whether alternatives to CPT such as TAX or DFT$_e$ would be able to provide a good account of choice data at large. After all, there are very few reported model fits of TAX to individual choice data (for two recent exceptions, see Glöckner, Hilbig, Henninger, & Fiedler, 2016; Scheibehenne and Pachur, 2015). To make matters worse, the reported performance of TAX relative to CPT have been somewhat inconsistent (see Glöckner et al., 2016; Scheibehenne & Pachur, 2015). In the case of DFT$_e$ the situation is even less clear given that no direct comparisons with either CPT or TAX have been reported.

6. Reanalysis of Rieskamp (2008)

We begin by fitting the different models to the choice data from Rieskamp (2008), in which participants were presented with binary lottery problems. The results from these analyses will allow us to compare model performance as well as their respective characterizations. Rieskamp's (2008, Study 2) data consists of choices to one-hundred and eighty option pairs involving two-outcome lotteries, some comprised of pure-gain (60 pairs), pure-loss (60), and mixed lotteries (60). Monetary outcomes ranged between -$100 and $100. These lotteries were randomly generated, with the constraint that within each lottery problem no option dominated the other. Thirty individuals took part in this study.

We fitted CPT, TAX, DFT$_e$, and EUT. The models were fitted to the 180 choices made by each individual participant. We also established different sets of parameters for lottery problems involving gain, loss, and mixed lotteries. The option of having distinct

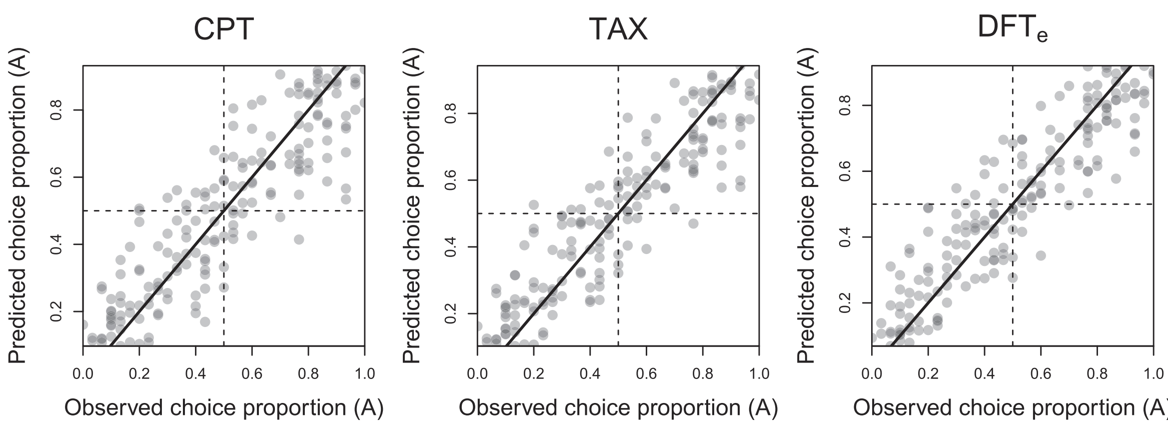

Fig. 7. Observed and predicted choice proportions for each lottery problem (aggregated across participants) in Rieskamp (2008). The predicted choice proportions associated with each model are based on each individual's median posterior parameters.

Table 2
Modeling of Rieskamp (2008): Mean Parameter Estimates (95% credibility intervals inside brackets).

Model	Parameter	Domain		
		Gain	Loss	Mixed
CPT	α	0.58 [0.52, 0.64]	0.73 [0.65, 0.81]	0.77 [0.70, 0.83]
	λ	—	—	1.19 [1.03, 1.36]
	γ	0.98 [0.89, 1.08]	0.97 [0.88, 1.06]	0.71 [0.63, 0.79]
	θ	1.37 [1.06, 1.70]	1.08 [0.78, 1.41]	0.51 [0.37, 0.69]
TAX	α	0.52 [0.46, 0.58]	0.61 [0.54, 0.69]	0.63 [0.57, 0.70]
	γ	0.93 [0.79, 1.07]	0.87 [0.75, 1.00]	0.62 [0.53, 0.72]
	δ	−0.05 [−0.23, 0.13]	-0.26 [-0.45, −0.07]	0.11 [-0.06, 0.27]
	θ	1.74 [1.39, 2.11]	1.40 [1.04, 1.77]	0.87 [0.65, 1.12]
DFT_e	α	0.74 [0.61, 0.86]	1.02 [0.90, 1.15]	0.82 [.73,.92]
	τ	—	—	0.03 [−0.12, 0.18]
	π	.68 [.62,.73]	.69 [.64,.74]	.63 [.57,.68]
	θ	1.91 [1.74, 2.10]	2.09 [1.90, 2.29]	2.72 [2.46, 2.96]

parameters for the latter case was based on previous results showing that the subjective representations observed in mixed lotteries can differ from the pure-gain and pure-loss cases (e.g., Wu & Markle, 2008; see also Footnote 7). Parameter estimation and model comparison were conducted within a Bayesian framework (Gelman et al., 2014; Lee & Wagenmakers, 2014). Succinctly, the uncertainty concerning each model parameter is represented by a probability distribution over the range of values that parameter can take on. Specifically, we establish a *prior distribution* for each parameter, which will then be updated by the observed data and the model's likelihood function using Bayes' Theorem, resulting in a *posterior distribution*. In the present case, we adopted non-informative uniform parameter priors. Model comparisons were conducted using Bayes Factors (Kass & Raftery, 1995). For further details on the Bayesian modeling, see the Appendix.

6.1. Modeling results and discussion

Fig. 7 illustrates the ability of the three candidate models to account for Rieskamp's (2008) data, by plotting the expected choice proportion associated with each lottery problem given each individual's posterior parameter distributions. The predicted response proportions generally follow the observed choice proportions, although with some degree of dispersion that can be largely attributed to the stochasticity usually found in participants' choices (for a discussion, see Loomes, 2005). Also of note is the fact that the range of predictions under which the data are most likely do not necessarily minimize the number of qualitative misses (i.e., points in the upper-left and lower-right quadrants in Fig. 7; see Glöckner and Pachur, 2012). In terms of goodness of fit, the three candidate models yield very similar fits, and all fit better than EUT. For reference, the rank-order correlations between the predicted and observed choice proportions were 0.89 (CPT), 0.91 (TAX), and 0.91 (DFT_e). In the case of EUT, the correlation was only 0.84. Each model's goodness of fit across individuals can be quantified via their mean posterior summed log-likelihoods (and 95% credibility intervals), which were $-2688[-2716, -2664]$, $-2617[-2644, -2593]$, and $-2616[-2641, -2595]$. EUT's log-likelihood was $-2881[-2903, -2862]$, a considerably worse result.

As shown in Fig. 6, in terms of log Bayes Factors, DFT_e outperformed both CPT and TAX for 83% of the participants, with median log Bayes Factors of 5.09 and 4.03, respectively. TAX slightly outperforms CPT (it wins for 56% of the participants). EUT was the best faring model for only one participant, a result indicating that the characterization of choices almost always requires a weighting of outcomes that deviates from their associated probabilities. The fact that CPT, TAX, and DFT_e yielded similar goodness-of-fit results but different Bayes Factors indicates that DFT_e's success is primarily due to it being a considerably more constrained model than either CPT and TAX.[12]

The mean posterior parameter estimates for each model are reported in Table 2. In the case of CPT, these estimates are in line with previous reports, but the weighting of probabilities was generally very close to linear with the exception of lottery problems involving mixed outcomes where it took on a more clear inverse-S-shape. Also, note that the mean λ was only 1.19, suggesting a near-absence of loss aversion as defined by CPT. The characterization obtained with TAX was somewhat different from the 'prior' parameters reported by Birnbaum and colleagues (e.g., Birnbaum, 2008). In the context of gain and mixed lottery problems, we do not observe a clear attention exchange from the best to the worst outcomes. The mean δ estimates associated with these domains (see Table 2) are at odds with the strong attention exchange that the 'prior' $\delta = 1$ assumes. One reason for this result is that the binary lotteries used in

[12] One concern surrounding the use of Bayes Factors in model comparisons is their sensitivity to the selection of parameter priors. As a robustness check, we also fitted the models using the maximum-likelihood method and compared them using the Akaike Information Criterion (AIC). The reason for choosing this model-selection criterion was the fact that it imposes a relatively lenient penalty on flexibility and does not require the specification of parameter priors. We found DFT_e outperforming CPT and TAX in 57% and 53% of the individual datasets, respectively. None of these differences was found to be statistically significant using a Wilcoxon test (smallest $p = .24$).

Rieskamp's (2008) study do not permit one to disentangle the role of attention exchange from the subjective representations of value and/or probabilities. In fact, the 'prior' $\delta = 1$ used by Birnbaum and colleagues was based on results from a previous study involving multiple outcomes (Birnbaum & McIntosh, 1996; Birnbaum & Stegner, 1979) as well as studies in which the utility scale was experimentally constrained (e.g., Birnbaum and Sutton, 1992).

In the case of DFT_e, the individual estimates are similar to the ones reported by Bhatia (2014), with participants generally engaging in non-proportional sampling with a probability of roughly .35. One interpretation of these estimates is that for about one third of their deliberation period, participants were considering the different outcomes of the lotteries without taking into account their respective probabilities. Analogous to parameter λ in CPT, the mean τ estimate obtained suggests that there is no asymmetry between gains and losses in the case of non-proportional sampling. When ignoring the outcomes' respective probabilities, participants paid attention to losses relative to gains only a slightly bit more.

Finally, we also asked whether TAX and DFT_e can predict the Allais' choice paradoxes as well as Birnbaum's strong choice paradoxes reported in Table 1. Using the individual mean parameter estimates obtained with Rieskamp's (2008) data, we found the answer to be positive on both counts: In the case of the Allais' Common Consequence Paradox, TAX predicts most people will choose $\mathcal{A}$ and $\mathcal{B}'$ (.54 and.95), and so does DFT_e (.73 and.56). Similar results are found in the case of Allais' Common Ratio Paradox, with most preferring $\mathcal{A}$ and $\mathcal{B}'$ (TAX:.62 and.48; DFT_e:.66 and.54). Turning to Birnbaum's paradoxes shown in Table 1, the expected choice proportions yielded by both TAX and DFT_e also capture the general pattern of choices.

The reanalysis of Rieskamp (2008) indicates that both TAX and DFT_e are able to fit participants' choices while also predicting the different choice paradoxes. These optimistic results will be put to the test in the experiment below, in which we present participants with a diverse set of lottery problems that, among other things, test the different choice paradoxes.

7. Experiment

7.1. Participants, procedure, and materials

Fifty-five participants (41 women, median age = 22 (SD = 6.49)) took part in a two-hour session at the Cognitive and Decision Sciences laboratory at the University of Basel, Switzerland. Each session had at most four participants, but the majority of them were conducted with a single participant. In exchange for their participation, they received course credit and a baseline payment of 5 Swiss Francs (CHF). Based on their choices, this payment was later updated to a value ranging between 0 and 10 CHF (for details, see below). The experiment was implemented in Psychopy (Peirce, 2007).

After being informed of the nature of the decision making task and trying out an example decision problem, participants encountered 313 decision problems with values ranging from −1000 to 1000 CHF, many involving binary lotteries that were used in previous studies that fitted CPT (e.g., Kellen et al., 2016).[13] One subset of lottery problems taken from Brooks et al. (2014) involved multiple-event lotteries and was designed to evaluate basic properties such as strong risk attitudes, and gain-loss asymmetries. Finally, we also included a number of lottery problems focused on testing different critical properties violated in Birnbaum's choice paradoxes: Stochastic dominance, event splitting and coalescing, upper-tail independence, and lower/upper cumulative independence. A list of these sets of lottery problems is provided in the Appendix. A complete list of all 313 lottery problems is provided in the Supplemental Materials at Open Science Framework (https://osf.io/ngc45/?view_only=befb58af7bc04f00a5afb201733aedf9). Unlike the earlier studies by Birnbaum, we did not rely on repeated lottery problems. Instead, we created new lottery problems by multiplying all outcomes by a common factor (1.5). These new trials can be seen as "scaled replications" given that, under the assumption that utilities follow a power function, preferences remain invariant to scaling by a common factor (for a proof, see Cho & Fisher, 2000).

When using decision problems that include multiple-event lotteries, there is the concern that participants might be overwhelmed with the amount of information provided. In order to ease the demands placed on these decision problems, the different options were presented with colored pie charts (see Fig. 8) that included both monetary outcomes and probability information. A color scheme was used to help participants to distinguish between the potential gains (green) and losses (red), as well as the rank order of the outcomes. More intense colors were used to represent more extreme outcomes, and events yielding the same monetary outcome had the same color. In order to discourage participants from responding quickly without evaluating the different options, they were required to select one of the options with the mouse (by pressing a button below their preferred option) and then confirming their choice using the 'OK' button. The right panel of Fig. 8 shows two lottery problems used to test event splitting and coalescing.

7.2. Results

7.2.1. Quality control

Given the large number of lottery pairs encountered by the participants, it is important to first and foremost check the overall quality of the data. We were able to evaluate the data quality by checking the response proportions in cases where one lottery transparently dominates the other. When confronted with the lottery pairs $\mathcal{A} = \begin{pmatrix} \$96 & \$14 & \$12 \\ .90 & .05 & .05 \end{pmatrix}$ and $\mathcal{B} = \begin{pmatrix} \$96 & \$90 & \$12 \\ .25 & .05 & .70 \end{pmatrix}$, none of the participants chose the objectively worse $\mathcal{B}$ option. For lottery pair $\mathcal{A} = \begin{pmatrix} \$12 & \$10 & \$0 & -\$10 \\ .20 & .20 & .40 & .20 \end{pmatrix}$, $\mathcal{B} = \begin{pmatrix} \$12 & \$5 & \$0 & -\$10 \\ .20 & .20 & .40 & .20 \end{pmatrix}$ only

[13] Due to the similar exchange rates between Swiss Francs and US dollars, we will keep using the '$' sign throughout this manuscript.

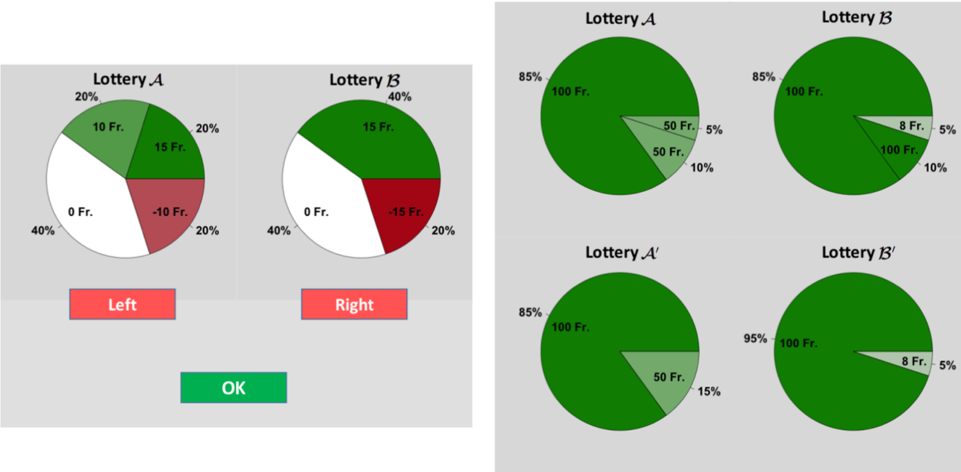

Fig. 8. Left Panel: Example of a decision problem. Right Panel: Two of the lottery pairs used to test event-splitting and coalescing.

7% chose the dominated option $\mathcal{B}$. These response patterns suggest that participants were generally attentive to the lottery problems.

7.2.2. Strong risk attitudes and gain-loss asymmetries

The evaluation of strong risk attitudes was based on a selection of lottery problems constructed by Brooks et al. (2014), in which two lotteries with the same expected value but different variances were compared. For example, the 'riskier' lottery $\mathcal{R} = \begin{pmatrix} \$15 & \$11 & \$2 & \$0 \\ .40 & .20 & .20 & .20 \end{pmatrix}$ and the 'safer' lottery $\mathcal{S} = \begin{pmatrix} \$15 & \$11 & \$10 & \$5 & \$2 \\ .20 & .20 & .20 & .20 & .20 \end{pmatrix}$. Strong risk-averse individuals are expected to choose the lottery with the lowest variance, whereas the risk-seeking individuals are expected to choose the higher-variance option. As shown in top row of Fig. 9, the choice proportions were in line with our expectations, and consistent with the notion that the subjective representation of monetary outcomes is well captured by a utility function that is convex for losses and concave for gains (Baucells & Heukamp, 2006; Brooks et al., 2014). This pattern held whether we aggregated choices across people or lotteries, although the latter allowed us to observe the presence of individual differences.

Moving on to the problems involving mixed lotteries with the same expected value but different variances, for example,

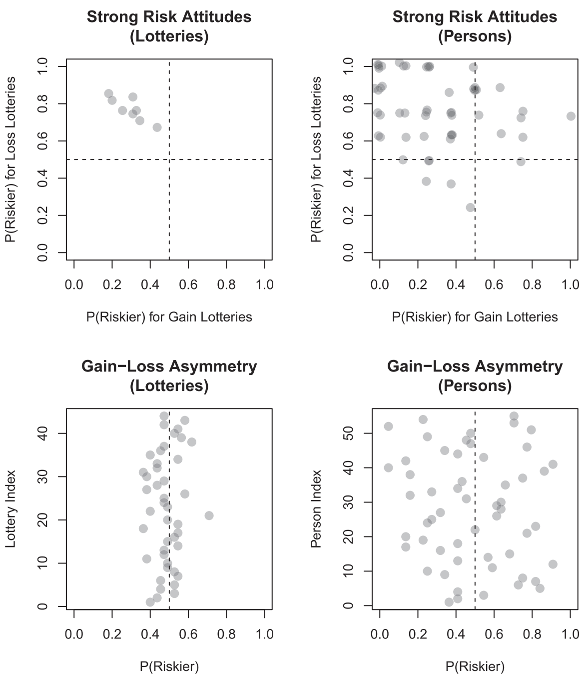

Fig. 9. Top Row: Proportion of riskier-lottery choices in pure-gain and pure-loss lottery problems (eight lottery problems and fifty-five participants). Bottom Row: Proportion of riskier-lottery choices in mixed lottery problems (forty-four lottery problems and fifty-five participants).

Table 3
Response proportions in a selection of lottery problems associated with choice paradoxes, and the respective model predictions.

				Predictions		
	Choice Paradox		$P(\mathcal{A})$	CPT	TAX	DFT_e
ESC	$\mathcal{A} = \begin{pmatrix} \$100 & \$50 \\ .85 & .15 \end{pmatrix}$	$\mathcal{B} = \begin{pmatrix} \$100 & \$8 \\ .95 & .05 \end{pmatrix}$	.62	.56	.62	.71
	$\mathcal{A}' = \begin{pmatrix} \$100 & \$50 & \$50 \\ .85 & .10 & .05 \end{pmatrix}$	$\mathcal{B}' = \begin{pmatrix} \$100 & \$100 & \$8 \\ .85 & .10 & .05 \end{pmatrix}$	.29	.56	.49	.52
SD	$\mathcal{A} = \begin{pmatrix} \$96 & \$14 & \$12 \\ .90 & .05 & .05 \end{pmatrix}$	$\mathcal{B} = \begin{pmatrix} \$96 & \$90 & \$12 \\ .85 & .05 & .10 \end{pmatrix}$	.22	.51	.34	.30
UTI	$\mathcal{A} = \begin{pmatrix} \$110 & \$44 & \$40 \\ .80 & .10 & .10 \end{pmatrix}$	$\mathcal{B} = \begin{pmatrix} \$110 & \$96 & \$10 \\ .80 & .10 & .10 \end{pmatrix}$	.29	.57	.48	.48
	$\mathcal{A}'' = \begin{pmatrix} \$96 & \$44 & \$40 \\ .80 & .10 & .10 \end{pmatrix}$	$\mathcal{B}'' = \begin{pmatrix} \$96 & \$10 \\ .90 & .10 \end{pmatrix}$	.69	.57	.57	.62
LCI	$\mathcal{A} = \begin{pmatrix} \$96 & \$12 & \$4 \\ .05 & .05 & .90 \end{pmatrix}$	$\mathcal{B} = \begin{pmatrix} \$52 & \$48 & \$4 \\ .10 & .10 & .80 \end{pmatrix}$	.53	.51	.47	.47
	$\mathcal{A}'' = \begin{pmatrix} \$96 & \$12 \\ .05 & .95 \end{pmatrix}$	$\mathcal{B}'' = \begin{pmatrix} \$52 & \$12 \\ .10 & .90 \end{pmatrix}$	.64	.51	.59	.68
UCI	$\mathcal{A} = \begin{pmatrix} \$110 & \$44 & \$40 \\ .80 & .10 & .10 \end{pmatrix}$	$\mathcal{B} = \begin{pmatrix} \$110 & \$98 & \$10 \\ .80 & .10 & .10 \end{pmatrix}$	.25	.57	.48	.48
	$\mathcal{A}'' = \begin{pmatrix} \$98 & \$40 \\ .80 & .20 \end{pmatrix}$	$\mathcal{B}'' = \begin{pmatrix} \$98 & \$10 \\ .90 & .10 \end{pmatrix}$	.64	.57	.62	.66

Note. ESC = Event splitting and coalescing, SD = First-order stochastic dominance, UTI = upper tail independence, LCI = lower-cumulative independence, UCI = upper cumulative independence. The values in column '$P(\mathcal{A})$' indicate the proportion of $\mathcal{A}$ choices made for the options in 'Choice Paradox'. Columns CPT, TAX, and DFT_e provide the expected proportion of $\mathcal{A}$ choice for each model, based on their mean posterior parameter estimates.

$\mathcal{R} = \begin{pmatrix} \$10 & \$8 & -\$8 & -\$10 & -\$12 \\ .20 & .20 & .20 & .20 & .20 \end{pmatrix}$ and the 'safer' lottery $\mathcal{S} = \begin{pmatrix} \$8 & \$5 & -\$5 & -\$8 & -\$12 \\ .20 & .20 & .20 & .20 & .20 \end{pmatrix}$, we fail to find clear-cut results. The bottom row of Fig. 9 shows that there was no clear preference for the safer, lower-variance option: In only 64% of the lottery pairs was the safer option the modal choice. Only 55% of the participants chose the safer options most often. These results suggest little to no gain-loss asymmetries in participants' preferences, in line with previous modeling work (Glöckner and Pachur, 2012; Kellen et al., 2016; Rieskamp, 2008; see also Ert & Erev, 2013).

7.2.3. Choice paradoxes

We now turn to the choice paradoxes. Here, we will only discuss the general pattern of results. A more careful model-based analysis is reported in the Appendix (section 'True-and-Error Modeling'). Table 3 reports the choice proportions observed in a representative selection of lottery pairs (e.g., the results for the pure-loss lotteries were pretty much mirror images). These proportions reveal choice patterns similar to the ones reviewed by Birnbaum (2008), indicating that the main paradoxes can be observed simultaneously, even when the relevant lottery problems are embedded in a large and diverse set.

Some of these results are particularly impressive given the way the lotteries were conveyed to the participants: As shown in Fig. 8, the differences between a pair of lottery problems used to test event splitting and coalescing are very subtle, making it very hard to dismiss such effects by simply claiming some misunderstanding on the participants' part. Also, the large number of lottery problems speaks against any explanation based on participants' inexperience or lack of references when dealing with lotteries. Moreover, these results were generally replicated in the 'scaled replication' trials. For example, consider the lottery pair in Table 3 testing stochastic dominance: In its respective scaled replication, 71% of the participants chose the dominated lottery $\mathcal{B} = \begin{pmatrix} \$144 & \$135 & \$18 \\ .85 & .05 & .10 \end{pmatrix}$ over $\mathcal{A} = \begin{pmatrix} \$144 & \$21 & \$18 \\ .90 & .05 & .05 \end{pmatrix}$. A total of 60% of the participants chose the dominated option in both lottery problems, and only 11% chose the dominating option twice.

7.3. Comparing CPT, TAX, DFT_e, and EUT

Fig. 10 illustrates the ability of each of the three theories to account for the choice data. Specifically, we contrast the predicted choice probabilities for each of the lottery problems based on the participants' mean posterior parameters against the observed choice proportions. In the top row, which focuses on the lottery problems not involved in the testing of paradoes, we see a pattern that is similar to the one observed with Rieskamp's (2008) data in Fig. 7. Model predictions generally tracked the observed choice proportions, although some dispersion and qualitative misses were found (see Glöckner and Pachur, 2012; Loomes, 2005). When turning to the lottery problems producing the choice paradoxes (bottom row), we see that CPT fares considerably worse than TAX and DFT_e,

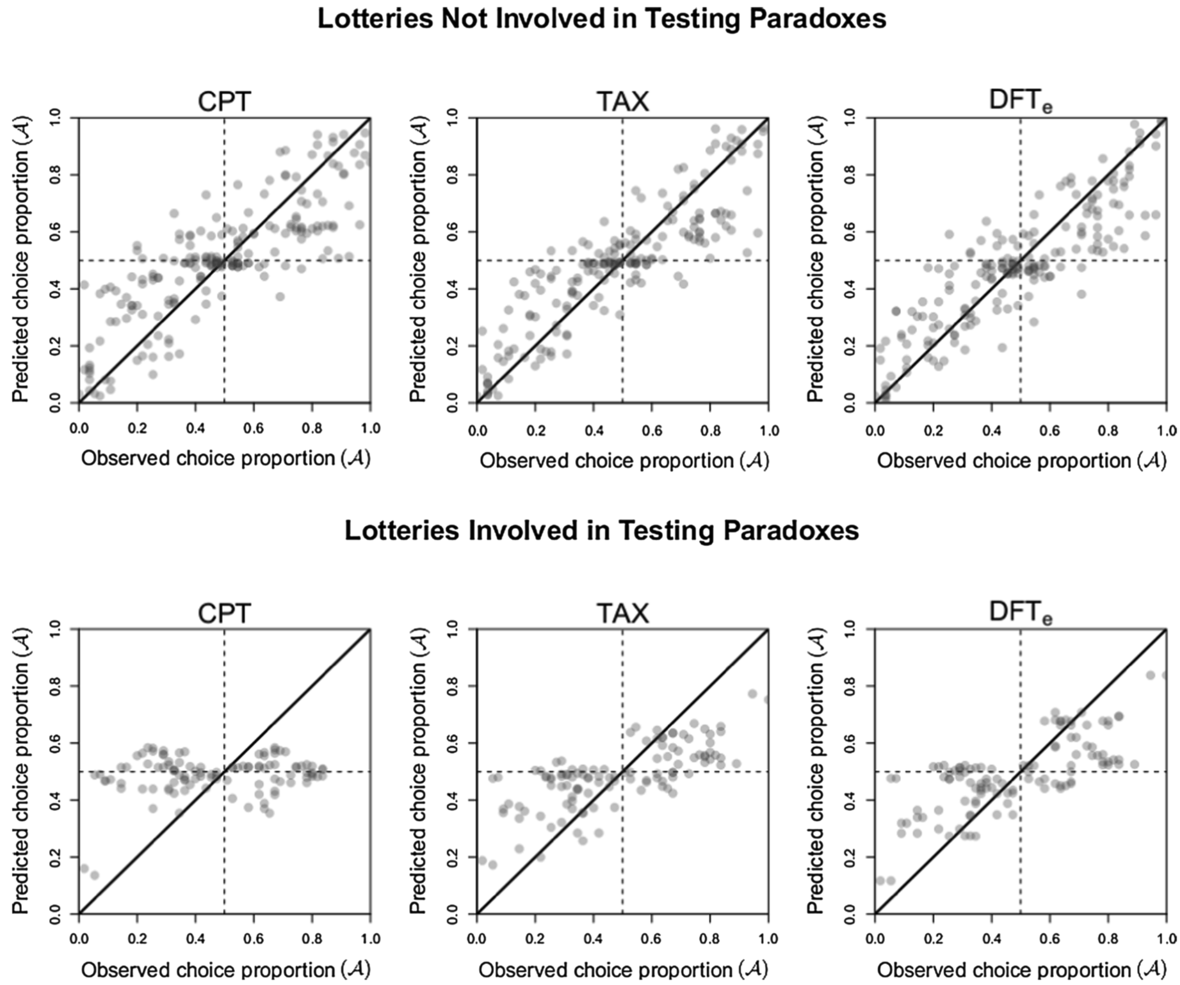

Fig. 10. Observed and predicted choice proportions for each lottery problem (aggregated across participants). The predicted choice proportions associated with each model are based on each participant's median posterior parameters. Top Row: All lottery problems. Bottom Row: Lottery problems used for testing choice paradoxes.

as many of its predictions lying very close to the .50 value, which corresponds to the model's best account of the observed choice 'paradoxes'. For instance, CPT is unable to capture the fact that 78% of the participants preferred $\begin{pmatrix} \$96 & \$90 & \$12 \\ .85 & .05 & .10 \end{pmatrix}$ over the stochastically dominating option $\begin{pmatrix} \$96 & \$14 & \$12 \\ .90 & .05 & .05 \end{pmatrix}$. The best CPT can do is predict that 51% of the participants preferred the latter option. The predictions associated with DFT$_e$ seem comparable to TAX: For reference, the rank-order correlations between observed and predicted choice proportions for CPT, TAX, and DFT$_e$ shown for all lottery problems (Fig. 10, top row) were 0.66, 0.83, and 0.81, respectively. For the lottery problems producing the choice paradoxes, they were 0.10, 0.69, and 0.67. These results are reflected in the selection of lottery problems in Table 3, where one can see how both TAX and DFT$_e$ capture the overall preference shifts, whereas CPT erroneously predicts preferences to remain invariant.

When conducting a more careful evaluation of goodness of fit (Birnbaum, 1973) by considering the posterior log-likelihoods for each model (summed across participants) we see that DFT$_e$ overall provided a better fit of the data: The posterior means were −10079 [-10119, −10043], −9551 [-9590, −9513], −9409 [-9443, −9375], and −10548 [-10577, −10522] for CPT, TAX, DFT$_e$, and EUT, respectively, with larger values indicating better fit.

The relative performance obtained from the summed goodness of fit values was corroborated by the log Bayes Factors obtained at the individual-participant level: First, none of the participants was better captured by EUT, once again corroborating the need for theoretical accounts that include some form of subjective probability and/or attention weighting. Overall, DFT$_e$ outperformed both competitors for 44 (80%) participants, with a median log Bayes Factor of 8.88. These differences indicate strong support for the winning model. TAX outperformed CPT for 49 (89%) participants, with a median log Bayes Factor of 6.21. Overall, DFT$_e$ provides a better account of the data at large, an account that does not hinge on unwarranted flexibility. If anything, DFT$_e$ is highly constrained

Table 4
Present Experiment: Mean Parameter Estimates (95% credibility intervals inside brackets).

Model	Parameter	Domain		
		Gain	Loss	Mixed
CPT	α	0.43 [0.40, 0.46]	0.45 [0.40, 0.50]	0.43 [0.38, 0.47]
	λ	—	—	1.34 [1.22, 1.46]
	γ	0.84 [0.79, 0.90]	1.00 [0.93, 1.07]	0.72 [0.65, 0.80]
	θ	1.46 [1.29, 1.62]	1.67 [1.46, 1.89]	1.06 [0.86, 1.26]
TAX	α	0.43 [0.40, 0.46]	0.41 [0.39, 0.44]	0.43 [0.39, 0.47]
	γ	0.66 [0.58, 0.74]	0.73 [0.65, 0.81]	0.69 [0.61, 0.78]
	δ	−0.12 [-0.22, −0.01]	−0.36 [-0.46, −0.26]	0.26 [0.15, 0.36]
	θ	1.96 [1.74, 2.18]	2.17 [1.94, 2.40]	1.04 [0.89, 1.20]
DFT_e	α	0.60 [0.56, 0.64]	0.70 [0.65, 0.74]	0.63 [0.57. 0.69]
	τ	—	—	−0.18 [-0.26, −0.10]
	π	.64 [.61,.67]	.70 [.67,.73]	.59 [.55,.64]
	θ	2.63 [2.49, 2.78]	2.70 [2.55, 2.84]	2.91 [2.72, 3.09]

in the way it can affect the weighting of events.[14] When inspecting the sample of lottery pairs reported in Table 3, we see that the predictions coming from both TAX and DFT_e can track the major differences in modal choices, despite some clear room for improvement.

Summaries of the mean posterior parameters of each model are given in Table 4. In the case of CPT, it is only worth noting that loss aversion parameter λ was on average slightly larger than 1, indicating that according to CPT, losses only loom a bit larger than gains and that the probability weighting functions were generally found to be inverse-S-shaped. Across all models, the utility functions were generally found to be concave and convex for gains and losses respectively (i.e., $\alpha < 1$). Across the different types of lottery pairs, the proportion of individual posterior mean α estimates below 1 was .98, 1, and .88 for CPT, TAX, and DFT_e, respectively.

In the case of TAX, it is interesting to note that once again there was no clear-cut direction for the attention transfer in pure-gain lotteries, with a mean posterior of $\delta = -0.12$. In the case of pure-loss and mixed lotteries, the mean δ parameters suggested a transfer from worst to best, and from best to worst, in line with previous characterizations (see Birnbaum & Bahra, 2007). As we will show later on, these results cannot be attributed to an inability to estimate TAX parameters using the present experimental design.

Regarding DFT_e, the parameter estimates show that a considerable portion of the samples (between one third and one half) are not proportional to the events' probabilities, leading to an overweighting of small probabilities and underweighting of large probabilities. The observed gain-loss asymmetries (as captured by parameter τ) reveal a slight overweighting of losses relative to gains in non-proportional sampling. Consider a lottery $\begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ .40 & .30 & .20 & .10 \end{pmatrix}$. In the domain of gains, the sampling probabilities of the four events were .35, .28, .22, and .15. In the case of losses, .36, .29, .21, and .14. Finally, in the case of mixed events, with x_1 and x_2 being gains, x_3 and x_4 losses, .32, 26, .24, and .18.

8. General discussion

In this work, we pitted CPT against two alternative models: TAX, a model that has been long associated with the aforementioned paradoxes (Birnbaum, 2008), and DFT_e (Bhatia, 2014). Both TAX and DFT_e offer alternatives to CPT's complex scheme of rank- and sign-dependent weighting using cumulative probabilities: Whereas TAX assumes that attention is transferred between better and worse outcomes (at least in the context of gains), DFT_e assumes proportional and non-proportional event sampling mechanisms, the latter sampling events with equal probability. Our model fits show that both TAX and DFT_e can account for participants' choices, with our Bayesian model selection indices preferring DFT_e due to its greater simplicity. This simplicity is manifested in the fact that DFT_e often produces weighting schemes that only slightly differ from EUT. In contrast, TAX allows for a wide range of attention-transfer schemes. Note however, that the present preference for DFT_e, which is to a large extent grounded on model flexibility, *does not* imply a dismissal of TAX. After all, the theory did not fail to account for the present choice data, and the present study only amounts to a sample of possible lottery problems and choice paradoxes. We will return to TAX later in this discussion.

The success of the DFT_e model shows that one can safely abandon CPT in favor of a model that is in many ways simpler. DFT_e's account of decisions under risk can be summarized as follows: *individuals choose by imagining the different possible events taking place with the described probabilities. During this deliberation period, individuals are sometimes distracted such that they think about the different events while ignoring their respective probabilities.* The simplicity of the sampling regime proposed by DFT_e implies that many different

[14] We also fitted the models to individual-participant data using the maximum-likelihood method and compared their performance using AIC. DFT_e outperformed both CPT (87%) and TAX (67%) for most participants, with performance differences that were found to be statistically significant in both cases (Wilcoxon test; largest $p = .007$). TAX also outperformed CPT (95%; $W = 1501$, $p < .001$).

avenues for model future development could be pursued. For example, preferences could be allowed to change across time (Bhatia & Loomes, 2017; Loomes, Moffatt, & Sugden, 2002; Marchiori, Di Guida, & Erev, 2015), or one could introduce the possibility of individuals sometimes relying on alternative sampling and/or comparison strategies (e.g., Gonzalez, Lerch, & Lebiere, 2003; Pachur et al., 2017). Also, one could change the non-proportional sampling so that it can be sensitive to certain aspects of the lotteries such as the rank order of events, best/worst case scenarios, or previous experiences among many other possibilities. Also relevant is the connection between the two sampling regimes and the interplay between affective and deliberative processes (e.g., Loewenstein, O'Donoghue, & Bhatia, 2015; Mukherjee, 2010). As reviewed by Loewenstein et al. (2015), there are physiological results which strongly suggest that affective responses are a function of the magnitude of the stimuli but not their probability of occurring (e.g., heart-rate increase being a function of the intensity of a likely future shock, but not its probability). The connection between affective processing and non-proportional sampling opens a wide range of hypotheses that go beyond decision-making under risk and include, among other things, intertemporal choices and social preferences.

8.1. Overall gain/loss probabilities in multiple-event lotteries

One distinctive aspect of the present study is that we did not restrict ourselves to binary lotteries. Luce (2000, p. 286) remarked that "*the issue of a suitable theory for gambles with three or more consequences is pretty much in the air*". One of the constraints theories face when attempting to characterize multiple-event lotteries is the fact that individuals' preferences are sensitive to the *overall probabilities* of gaining and/or losing money (e.g., Lopes & Oden, 1999; Payne, 2005; Payne, Laughhunn, & Crum, 1980). For instance, Payne et al. (1980) demonstrated how a constant shift across all events can reverse preferences: Most participants (83%) prefer $\mathcal{B}^+ = \begin{pmatrix} \$40 & \$30 & \$15 \\ .30 & .50 & .20 \end{pmatrix}$ to $\mathcal{A}^+ = \begin{pmatrix} \$74 & \$30 & -\$25 \\ .50 & .10 & .40 \end{pmatrix}$, but when shifting all events by $-\$60$, most (67%) preferred the resulting $\mathcal{A}^- = \begin{pmatrix} \$14 & -\$30 & -\$85 \\ .50 & .10 & .40 \end{pmatrix}$ over $\mathcal{B}^- = \begin{pmatrix} -\$20 & -\$30 & -\$45 \\ .30 & .50 & 20 \end{pmatrix}$. This result is expected under DFT$_e$: When using the average parameters obtained with mixed lotteries, $\alpha = 0.63$, $\pi = .59$, and $\tau = -0.18$, we see that $V(\mathcal{B}^+) = 8.29 > V(\mathcal{A}^+) = 4.66$, and $V(\mathcal{A}^-) = -6.06 > V(\mathcal{B}^-) = -8.55$. More recently, Payne (2005) showed that participants prefer improving the overall probability of winning than improving the best outcome in a lottery. Participants were first given a lottery $C = \begin{pmatrix} \$100 & \$50 & \$0 & -\$25 & -\$50 \\ .20 & .20 & .20 & .20 & .20 \end{pmatrix}$, and gave the possibility of improving it either by adding \$38 to the event yielding \$100 or by adding it to event yielding \$0. Most participants preferred the second option, increasing the overall probability of a gain. Again, this result is expected under DFT$_e$, as the subjective values of $C' = \begin{pmatrix} \$138 & \$50 & \$0 & -\$25 & -\$50 \\ .20 & .20 & .20 & .20 & .20 \end{pmatrix}$ and $C'' = \begin{pmatrix} \$100 & \$50 & \$38 & -\$25 & -\$50 \\ .20 & .20 & .20 & .20 & .20 \end{pmatrix}$ are 1.72 and 2.75, respectively.

8.2. Non-proportional sampling and contextual effects

The discussion of individuals' sensitivity to the possibility of gains and losses brings us to the specific way we extended DFT$_e$ to accommodate gain-loss asymmetries. The motivation for extending the model through its non-proportional sampling, rather than simply introducing a λ parameter in the utility function as done in CPT, was the fact that it can accommodate phenomena that would be problematic otherwise. For example, Walasek and Stewart (2015) showed that 'losses loom larger than gains', with people generally refusing to play mixed lotteries when lotteries tended to be attractive (i.e., with positive expected values). When the mixed lotteries were generally unattractive, the observed choices suggested that 'gains loom larger than losses' (i.e., $\lambda < 1$).

This demonstration of context dependency is problematic for CPT when attempting to capture them via the loss aversion parameter λ as it precludes the assumption of a stable subjective representation of monetary outcomes (see also Ert & Erev, 2013). Similarly problematic findings were previously reported by Lerner, Small, and Loewenstein (2004), who showed that the endowment effect (greater selling versus buying prices), reverses when individuals' negative emotional state prompts them to change their current status quo. What the present extension of DFT$_e$ offers is a way to accommodate these context dependencies by arguing that, depending on the attractiveness of the lotteries (or the decision maker's emotional state, goals), individuals' non-proportional sampling might be biased towards/against sampling gains relative to losses. In other words, when individuals focus on the events exclusively (i.e., ignoring their probabilities), they might not pay the same attention to the gains and losses. The relative attention would then depend on the relative attractiveness of the lotteries such that more attention is given to these losses when they are attractive, but to the gains when they are not. The same can be said when contrasting different perspectives (e.g., buying versus selling). Note that this differential-attention account bears some similarities with Busemeyer and Townsend (1993) incorporation of approach-avoidance conflicts in DFT, conflicts which can also be context dependent (e.g., dependent on reference points, goals and/or aspirations; see also Lopes & Oden, 1999; Wang & Johnson, 2012).

8.3. The role and estimability of δ in TAX

The δ estimates obtained in the present study suggest that the attention-transfer process postulated by TAX, at least in the domain of gains, plays little to no role in describing people's choices in this experiment. This result is at odds with the prior $\delta = 1$ (for gains) that is commonly adopted. As previously discussed, when looking back at the origin of the 'prior' parameters, we see that they are based on a variety of previous findings across different studies, including 'scale-convergence' constraints on the utility function (e.g., Birnbaum and Sutton, 1992) as well as *violations of restricted branch independence* (Birnbaum & Chavez, 1997: Birnbaum & McIntosh, 1996; Birnbaum and Stegner, 1996). Consider the following lottery problem in which both lotteries have a common branch with outcome z:

$$\mathcal{A} = \begin{pmatrix} z & \$44 & \$40 \\ .50 & .25 & .25 \end{pmatrix} \quad \mathcal{B} = \begin{pmatrix} z & \$98 & \$10 \\ .50 & .25 & .25 \end{pmatrix}$$

According to EUT, preferences for $\mathcal{A}/\mathcal{B}$ is invariant to manipulations of the common branch z. This property is known as restricted branch independence, and multiple studies have shown it to fail: For example, Birnbaum and Chavez showed that most participants prefer $\mathcal{A}$ when $z = \$5$, but most also prefer $\mathcal{B}$ when $z = \$111$. CPT with an inverse-S-shaped probability-weighting function expects violations,*but in the opposite direction*: A preference for $\mathcal{B}$ when $z = \$5$ and $\mathcal{A}$ when $z = \$111$. TAX with prior parameters produces the observed pattern. However, this pattern is *not* predicted under the parameter estimates obtained in the present study. This failure can be attributed to δ values that are too low.

One possible explanation for our discrepant results is that our selection of lottery problems does not impose the same type of constraints that some of these previous studies did. For instance, we did not test for violations of restricted branch independence, nor did we include manipulations that constrain the scale of the utility function. Given these differences, it is possible that our study is unable to accurately characterize the attention exchange quantified by δ. [15] One way to assess this possibility is to evaluate the ability to recover TAX parameters under the present selection of lottery problems. We simulated choice data using TAX and different sets of parameter values. The parameters estimated from these artificial data were compared with the original data-generating values. If our selection of lottery problems is indeed incapable of providing an accurate TAX characterization, then the resulting parameter estimates should not resemble the data-generating parameter values in any way. To make things simple, we will only consider the 109 pure-gain lottery pairs and estimate parameters using the maximum-likelihood method.

We conducted three different simulations: In the first one, we generated data using the maximum-likelihood parameter estimates obtained when fitting each person's choice data from the present study. These estimates should generate data that resembles the choices that we observed. In the second simulation, we first sampled parameter values (fifty-five sets of values, to match the number of persons in our study) from ranges that are consistent with the aforementioned TAX 'priors'.[16] Because we obtained similar results in both simulations, we will focus on the second one. The results shown in Fig. 11 show that parameters were generally well recovered (these results were stable across multiple simulation runs), with the exception of ϕ. For reference, these correlations are quite good when compared with previous results comparing individual parameter estimates across sessions (see Broomell and Bhatia, 2014; Glöckner and Pachur, 2012; Kellen et al., 2016; Zeisberger, Vrecko, & Langer, 2012). Also, note that ***we observe no underestimation of*** δ — if anything, the results suggest the contrary.

In our third simulation, we generated data using parameter values sampled from the aforementioned uniformed distributions, with the exception of δ, which was fixed to a constant value of -1, 0.5, 0, 0.5, or 1. The goal here was to evaluate the range of values taken by the δ estimates. The median δ estimates (and 2.5% and 97.5% quantiles) obtained were, in order, -0.92 [-1.50, -0.20], -0.43 [-1.02 -0.09], 0.01 [-0.30 0.27], 0.51 [0.27 0.85], and 1.01 [0.75 1.31]. The median δ estimates generally matched the respective data-generating values, and the overlap between the different estimate distributions was relatively minor, such that δ values between 0.5 and 1 were very unlikely to be estimated as being close to zero or below.

Altogether, the simulation results show that the set of lottery problems used in the present study introduces enough constraints for TAX parameters to be reasonably estimated, including δ. In other words, the discrepancy between the estimation approach used here and the ones previously used by Birnbaum and colleagues cannot be attributed to some deficiency of the former. Also, note that we were able to replicate a large number of behavioral phenomena, which speaks against the idea that there is something odd with the present experimental design and/or the data coming from it.

8.4. Incorporating violations of restricted branch independence in DFT_e

The assumptions underlying DFT_e imply restricted branch independence, such that preferences should not be affected by the value taken by the common outcome z.[17] This means that DFT_e in its current form cannot accommodate the rejections of restricted branch independence that have been reported in the literature (for an overview, see Birnbaum, 2008). Although this situation is unfortunate, it does not mean that a revised version of the model could not overcome this challenge. We briefly sketch two possibilities below.

One way of revising DFT_e consists of allowing for different rates of non-proportional sampling to 'better' and 'worse' outcomes, analogous to the differential sampling of gains and losses currently implemented (captured by a parameter τ). For example, if we split between better and worse outcomes based on their median value, then violations of restricted branch independence like the one reported above can be expected: With $\alpha = .60$, $\pi = .64$, and $\tau = -0.2$ (indicating a slight non-proportional oversampling of the 'worse' values), $\mathcal{A} = \begin{pmatrix} \$111 & \$44 & \$40 \\ .50 & .25 & .25 \end{pmatrix} = 12.44$, which is smaller than $\mathcal{B} = \begin{pmatrix} \$111 & \$98 & \$10 \\ .50 & .25 & .25 \end{pmatrix} = 12.67$, whereas $\mathcal{A}' = \begin{pmatrix} \$44 & \$40 & \$5 \\ .25 & .25 & .50 \end{pmatrix} = 6.31$ is

[15] As a sanity check we fitted a restricted version of TAX in which $u(x) = x$ to the pure-gain lotteries with outcomes no greater than \$150. The median maximum likelihood estimates for γ and δ were 0.64 and 0.19, respectively. Note that these results are similar to the ones reported by Scheibehenne and Pachur (2015), 0.64 and 0.33, respectively.

[16] Parameters were sampled from independent uniform distributions with the following ranges: α: [0.70, 1], γ: [0.50, 0.90], δ: [0.70, 1.30], ϕ: [0.50, 1.50].

[17] Strictly speaking, DFT_e expects choice probabilities to be affected by changes in the common branches, as the variance of the lotteries changes (see Eq. (15)). However, these differences do not lead to a switch of the most probable choice.

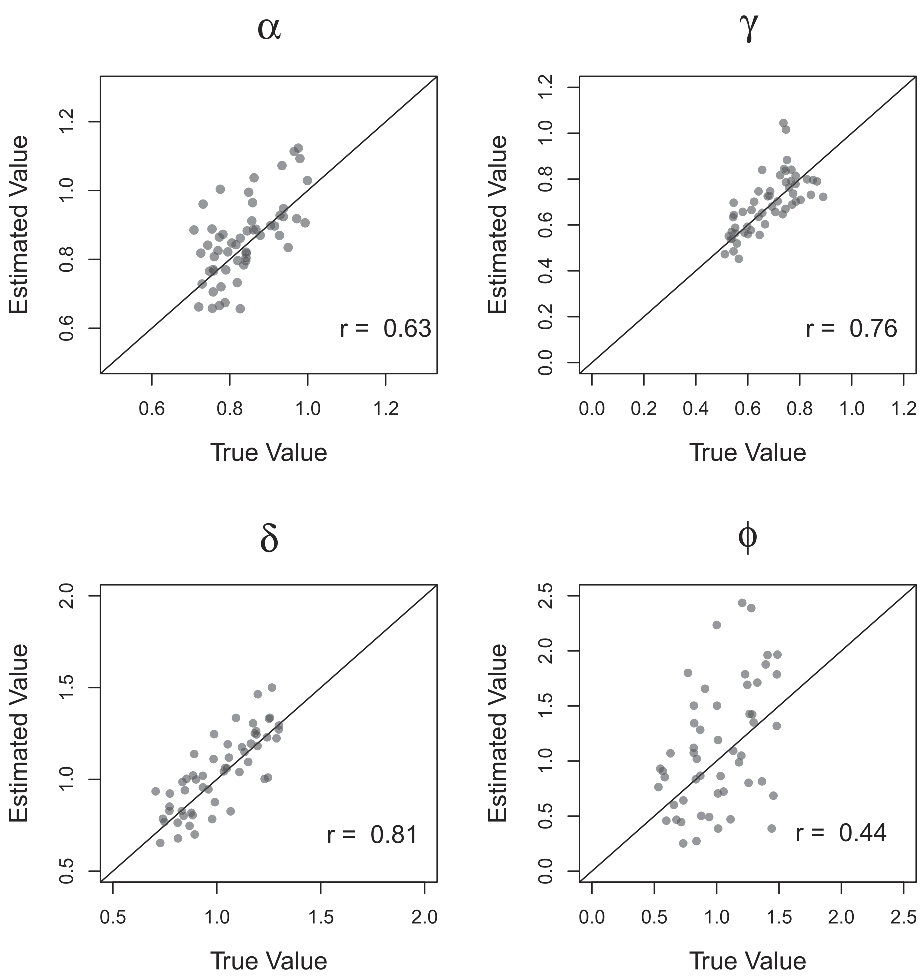

Fig. 11. True and estimated TAX parameters for 55 simulated participants (and their rank correlation).

greater than $\mathscr{B}' = \begin{pmatrix} \$98 & \$10 & \$5 \\ .25 & .25 & .50 \end{pmatrix} = 6.28$.

Another possibility involves the incorporation of a more involved attention-allocation process like the one recently proposed by Johnson and Busemeyer (2016) for DFT. We will only provide a streamline description of this account. Consider a lottery $\mathcal{L} = \begin{pmatrix} x_1 & x_2 & x_3 \\ p_1 & p_2 & p_3 \end{pmatrix}$. In a given 'mental play', the decision maker begins by focusing her attention on outcome x_i with probability μ_i. When the focus is on a given outcome, it is simulated with probability p_i. With probability $(1 - p_i)\eta$, there is no simulation but the outcome x_i stays in the focus of attention (η is understood as a '*dwell rate*'). Attention is shifted from an extreme outcome (x_1 or x_3) to the adjacent outcome (x_2) with probability $(1 - p_i)(1 - \eta)$, and from an non-extreme outcome (x_2) to adjacent one (x_1 or x_3) with probability $(1 - p_i)(1 - \eta)\frac{1}{2}$. When attention is shifted to an adjacent outcome, this outcome can be simulated with its respective probability. Fig. 12 illustrates this attention-allocation process.

The account proposed by Johnson and Busemeyer (2016) can produce violations of stochastic dominance and restricted branch independence but, as discussed by the authors, fails with other paradoxes such as event-splitting and coalescing or upper-tail independence. However, their proposal can be enriched by assuming that the above-described attention allocation process occurs with probability π, whereas any of the K outcomes can be directly simulated with probability $(1 - \pi)\frac{1}{K}$. As a first step, we evaluated the ability of this extended DFT model to expect all the choice paradoxes including restricted branch independence by varying parameters α, π, and η. We considered different starting values μ: *a)* Set $\mu_K = 1$, *b)* set all $\mu_k = \frac{1}{K}$, and *c)* set all $\mu_k = p_k$. The results shown in Fig. 13 indicate that all paradoxes are expected under parameter values that are not very far from the ones obtained with DFT$_e$. Also of note is the fact that all paradoxes appear to be compatible with a large range of 'dwell rates' (η).

Any revision of DFT$_e$ is likely to affect its performance relative to TAX, which is already able to account for violations of restricted branch independence in its current form. This possibility does not compromise the present results given that they did not imply a dismissal of TAX, only that it did not provide the most parsimonious account of the present body of behavioral phenomena. Future work should focus on identifying and testing critical properties distinguishing TAX and (revised) DFT$_e$. However, note that in order to

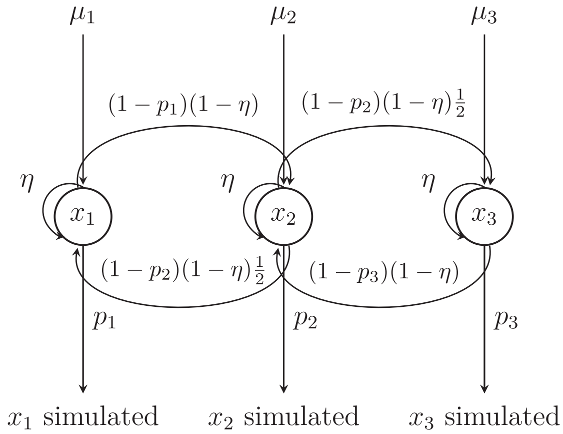

Fig. 12. Simplified schematic of the attention-allocation process proposed Johnson and Busemeyer (2016).

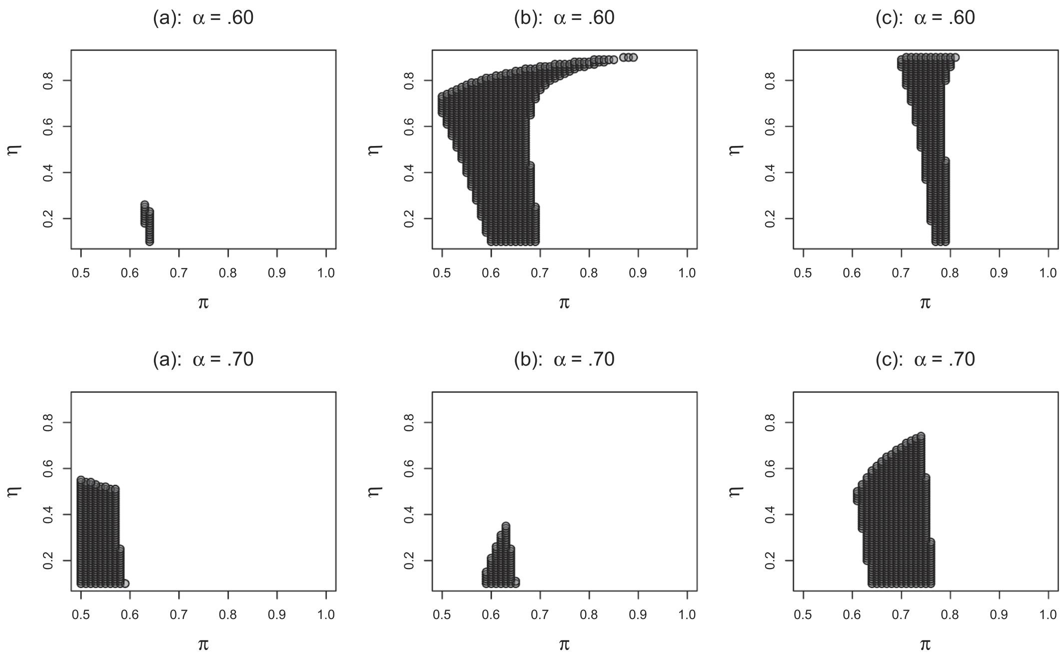

Fig. 13. Parameter values (with η = [.10, .90] and π = [.50, 1]), under different α values (.60 and.70) and μ constraints a, b, and (c) under which a DFT$_e$ model including Johnson and Busemeyer's (2016) attention-allocation process expects the examples of Birnbaum's strong paradoxes reported in Table 1 as well as the violation of restricted branch independence described in the body of text.

identify these critical properties, one also needs to clarify the actual role of attention exchange in TAX. In light of the present results, we first need to understand the extent to which the value of δ (not its estimation) is affected by the experimental design (for a related discussion, see Spektor, Kellen, and Hotaling, 2018). [18]

8.5. DFT$_e$'s relationship with other sampling-based models

As previously discussed, DFT assumes a sequential-sampling process that terminates when a boundary is reached (see Fig. 3). This type of characterization that can be found in many different domains such as memory and perception (e.g., Ratcliff & Smith, 2004; Starns, Ratcliff, & McKoon, 2012). A common aspect in all of these accounts is the presence of non-systematic stochastic noise or errors that can introduce systematic effects on behavior (e.g., Ratcliff and Smith, 2004; see also Bhatia and Loomes, 2017; Blavatskyy,

[18] One reviewer asked whether DFT$_e$ or any of its extensions are special cases of the '*general*' TAX model described by Eq. (7). We believe so, but in a way that is theoretically vacuous. To the best of our knowledge, the general TAX model has no testable properties (see Luce & Marley, 2005; Marley & Luce, 2005). Any testable model, not just DFT$_e$, can always be framed as a special case of a more general, untestable model.

2008). These effects can have very different accounts under theories that do not postulate such a noisy sampling processes. For instance, CPT argues that many phenomena such as the Allais' paradoxes are due to a distorted representation of probabilities that takes on an inverse-S shape. In the case of DFT_e, no such function is assumed. Instead, the model merely assumes the presence of noise or errors in the form of a non-proportional sampling process. The end result is an under/overweighting of certain events, yielding preferences that are consistent with inverse-S-shape representation of probabilities. A similar case has been recently made in the domain of probability judgments: Hilbert (2012) and Costello and Watts (2014), Costello and Watts (2016) showed that sampling errors could explain a considerably large body of behavioral phenomena that until then had been ascribed to people's reliance on heuristics or subjective representations that violate the laws of probability.

DFT_e has some similarities with other sampling-based models of decision-making under risk: For instance, Lin, Donkin, and Newell (2015) recently proposed a *Exemplar-Confusion model* that is formally very similar to DFT_e. According to this model, the sampled events can be incorrectly identified as one of the other events with some probability. The Exemplar-Confusion model has been developed with the intent of describing differences between choices involving *described* versus *experienced* options — *description-experience gaps* (Hertwig, Barron, Weber, and Erev, 2004; see also Kellen et al., 2016). A key aspect of the Exemplar-Confusion model is that the sampling process in experience-based decisions is not based on a uniform sampling; instead the sampling under error is proportional to the events sampled so far, which leads to an underweighting of rare events rather than an overweighting. One concern with this sampling assumption is that the underweighting of small probabilities that falls from it might not be empirically supported. Glöckner et al. (2016), and Kellen et al. (2016) showed that the weighting of experience-based decisions, after controlling for participants' experiences, still shows the typical pattern of overweighting of small probabilities. Kellen et al. compared description and experience-based choices in a within-subjects design and found that in the latter condition, individuals were more sensitive to differences between outcomes (higher α) and less sensitive to differences between probabilities (lower γ). In principle, these differences could be accommodated under DFT_e by simply having participants engage in a greater proportion of non-proportional sampling, which would amount to a greater focus on the events, irrespective of their probabilities. In addition to this concern, it is worth noting that the exemplar-confusion model's sampling process expects no event splitting/coalescing effects in experience-based decision making. Because the sampling is proportional to the experienced events, there is no possibility of oversampling the split events. To the best of our knowledge, the only study testing the effect of event splitting in experience-based decisions was reported by Erev et al. (2017), where no effect was found. But the informativeness of this result is questionable, as they also failed to observe an event-splitting effect when the lotteries were described (as done in the present study).

Another relevant model is the *Utility-Weighted Sampling model* recently proposed by Lieder, Griffiths, and Hsu (2017), which assumes choices are based on a sampling process that overweights extreme outcomes: when facing the options $\mathcal{A} = \begin{pmatrix} x_1 & x_2 & x_3 \\ p_1 & p_2 & p_3 \end{pmatrix}$ and $\mathcal{B} = \begin{pmatrix} y_1 & y_2 & y_3 \\ q_1 & q_2 & q_3 \end{pmatrix}$, individuals sample utility differences $\Delta U_{i,j} = u(x_i) - u(y_j)$ proportional to $p_i \times q_j \times |\Delta U_{i,j}|$. Although this model is able to account for a wide range of phenomena in both choice and probability judgments (see their Table A1), it cannot account for certain paradoxes such as violations of event-splitting and coalescing or violations of stochastic dominance, among others. Given the success of the Utility-Weighted Sampling model in accounting for many different types of judgments and effects, these failures can be seen as somewhat minor shortcomings. However, they raise the possibility that the sampling regime used depends on the nature of the judgments being made (e.g., probability judgments versus choices) or the nature of the information available (e.g., lotteries being described versus experienced). Future work should focus on better understanding the conditions under which different types of sampling are necessary and/or sufficient.

9. Conclusion

Recent modeling efforts in decision making have highlighted the importance of considering the unfolding of processes that lead to choices (e.g., Bhatia, 2014; Lieder et al., 2017). The present work shows that many behavioral phenomena, including the strong 'paradoxes of choice' reported by Birnbaum and colleagues, can be accounted for by a sequential-sampling model that proposes a very simple sampling regime. Our hope is that the success of this model contributes to a paradigm shift that is long overdue, away from Cumulative Prospect Theory and towards process-based accounts, like the ones within the realm of Decision Field Theory.

Appendix A

A.1. Bayesian estimation and model selection

Posterior samples were obtained using a differential evolution Markov Chain Monte Carlo algorithm (ter Braak & Vrugt, 2008), implemented in the `R` package `BayesianTools` (Hartig, Minunno, Paul, Cameron, & Ott, 2017). When comparing the ability of multiple models to account for the observed data (e.g., by computing the data's likelihood under each model), it is critical that their relative flexibility is taken into account, such that the most successful candidate model is the one providing the best balance between goodness of fit and parsimony (Kellen, Klauer, & Bröder, 2013; Myung, Cavagnaro, & Pitt, 2017). We compared models using *Bayes Factor* (BF; Kass & Raftery, 1995), a measure that quantifies the change in the relative evidence for two models/hypotheses that is obtained from the data. Let $f(\mathbf{X}, \mathfrak{M}_i, \Theta_{\mathfrak{M}_i})$ denote the likelihood of data vector $\mathbf{X}$ under model $\mathfrak{M}_i$ with parameter vector $\Theta_{\mathfrak{M}_i}$. For any two models $\mathfrak{M}_i$ and $\mathfrak{M}_j$, the Bayes Factor $BF_{i,j}$ is given by

$$BF_{i,j} = \frac{\int_{\Theta_{\mathfrak{M}_i}} f(\mathbf{X}, \mathfrak{M}_i, \Theta_{\mathfrak{M}_i})\, d\Theta_{\mathfrak{M}_i}}{\int_{\Theta_{\mathfrak{M}_j}} f(\mathbf{X}, \mathfrak{M}_j, \Theta_{\mathfrak{M}_j})\, d\Theta_{\mathfrak{M}_j}}. \tag{14}$$

The numerator and denominator correspond to the two models' respective marginal likelihoods — the integral of the model's likelihood function across the supported range of parameter values $\Theta_{\mathfrak{M}}$. Whereas typical goodness-of-fit measures rely on the maximum likelihood of the data under a given model, marginal likelihoods effectively compute a weighted-average likelihood where the weights correspond to the prior density given to each parameter value. Models with unwarranted flexibility are automatically penalized because they include predictions under which the data are very unlikely. Bayes factors larger than 1 indicate support for model $\mathfrak{M}_i$ relative to $\mathfrak{M}_j$, and values between 0 and 1 favor $\mathfrak{M}_j$ relative to $\mathfrak{M}_i$ (a value of 1 indicates that the evidence is equal). In order to establish a symmetrical scale for both models, it is convenient to consider the natural logarithm of the Bayes factor, in which positive and negative values favor $\mathfrak{M}_i$ and $\mathfrak{M}_j$, respectively, with values more extreme than ± 3.40 corresponding to very strong relative support. Marginal likelihoods were computed using the bridge-sampling algorithm (Meng & Wong, 1996; Meng & Schilling, 2002), implemented in the `R` package `bridgesampling` (Gronau et al., 2017).

A.2. Fitting CPT, TAX, DFT_e (and EUT)

The models were fitted separately to each participant's set of choices. We also established different sets of parameters for lottery problems involving gain, loss, and mixed lotteries. The option of having distinct parameters for the latter case was based on previous results showing that the subjective representations observed in mixed lotteries can differ from the pure-gain and pure-loss cases (e.g., Wu and Markle, 2008; see also Footnote 7).

We established uniform parameter priors that were non-informative: In terms of the utility function, prior α followed a uniform distribution between 0.01 and 2. Also, the prior for the choice-sensitivity parameter θ was set to range between 0.01 and 4. In the case of CPT, λ was assigned a prior ranging from 0.01 to 3. Also in CPT, the probability-weighting parameter γ was set to range between 0.28 and 1.72, the lower boundary being set by the function's monotonicity constraints (see Footnote 4). EUT is a special case of CPT in which γ and λ are both restricted to 1. Note that the average prior parameters of CPT yield predictions that are in line with EUT, at least in the domain of gains and losses. In the case of lottery problems with outcomes, the average prior λ is 1.5, which introduces a small degree of loss aversion. In total, CPT has ten parameters per person: three $\{\alpha, \gamma, \theta\}$ sets for pure-gain, pure-loss, and mixed lottery problems, and one parameter λ for the latter lottery problems.

In the case of TAX, we established priors for δ and γ that ranged from -1.5 to 1.5, and from 0.01 to 2, respectively. Finally, for DFT_e the priors for π and τ were set to range between 0 and 1, and between -1 and 1, respectively. The average prior parameters of TAX produce predictions that are in line with EUT.

DFT_e also has ten parameters per participant: Three $\{\alpha, \pi, \theta\}$ sets and one parameter τ for the mixed lottery problems. TAX has twelve parameters, namely three $\{\alpha, \gamma, \delta, \theta\}$ sets. Finally, EUT has three sets of two parameters $\{\alpha, \theta\}$, for a total of six parameters. In the case of DFT_e, the average prior parameters do not produce predictions in line with EUT because this model only makes such predictions when π is at its upper boundary of 1.

A.3. Strong risk attitudes and gain-loss asymmetries

Table A1
Table A2

Table A1
Set of Lottery Problems Testing Strong Risk Atttitudes (Gain domain).

	Lottery $\mathcal{A}$										Lottery $\mathcal{B}$									
Problem	x_1	x_2	x_3	x_4	x_5	p_1	p_2	p_3	p_4	p_5	x_1	x_2	x_3	x_4	x_5	p_1	p_2	p_3	p_4	p_5
1	\$10	\$5	\$0			.40	.20	.40			\$10	\$5	\$0			.20	.60	.20		
2	\$5	\$3	\$0			.40	.20	.40			\$5	\$3	\$2	\$0		.20	.40	.20	.20	
3	\$15	\$5	\$0			.40	.40	.20			\$15	\$10	\$5			.20	.20	.60		
4	\$10	\$5	\$0			.40	.40	.20			\$10	\$5				.20	.80			
5	\$15	\$5	\$0			.60	.20	.20			\$15	\$10	\$5			.40	.20	.40		
6	\$10	\$5	\$0			.60	.20	.20			\$10	\$5				.40	.60			
7	\$15	\$11	\$2	\$0		.40	.20	.20	.20		\$15	\$11	\$10	\$5	\$2	.20	.20	.20	.20	.20
8	\$11	\$10	\$6	\$2	\$0	.20	.20	.20	.20	.20	\$11	\$6	\$5	\$2		.20	.20	.40	.20	

Note. The lottery problems used in the loss domain correspond to the ones used in the gain domain, but with all outcomes multiplied by − 1.

Table A2
Set of Lottery Problems Testing Gain-Loss Asymmetries.

	Lottery $\mathcal{A}$										Lottery $\mathcal{B}$									
Problem	x_1	x_2	x_3	x_4	x_5	p_1	p_2	p_3	p_4	p_5	x_1	x_2	x_3	x_4	x_5	p_1	p_2	p_3	p_4	p_5
1	$15	$0	−$12	−$15		.20	.20	.40	.20		$10	$0	−$10	−$12		.20	.20	.20	.40	
2	$10	$0	−$8	−$10		.20	.20	.40	.20		$5	$0	−$5	−$8		.20	.20	.20	.40	
3	$5	$0	−$4	−$5		.20	.20	.40	.20		$2	$0	−$2	−$4		.20	.20	.20	.40	
4	$15	$12	$0	−$15		.20	.40	.20	.20		$12	$10		−$10		.40	.20	.20	.20	
5	$10	$8	$0	−$10		.20	.40	.20	.20		$8	$5		−$5		.40	.20	.20	.20	
6	$5	$4	$0	−$5		.20	.40	.20	.20		$4	$2		−$2		.40	.20	.20	.20	
7	$15	$12	−$12	−$15		.40	.20	.20	.20		$15	$12	$10	−$10	−$12	.20	.20	.20	.20	.20
8	$12	$10	$8	−$8	−$10	.20	.20	.20	.20	.20	$12	$8	$5	−$5	−$8	.20	.20	.20	.20	.20
9	$8	$5	$4	−$4	−$5	.20	.20	.20	.20	.20	$8	$4	$2	−$2	−$4	.20	.20	.20	.20	.20
10	$15	$12	−$12	−$15		.20	.20	.20	.40		$12	$10	−$10	−$12	−$15	.20	.20	.20	.20	.20
11	$10	$8	−$8	−$10	−$12	.20	.20	.20	.20	.20	$8	$5	−$5	−$8	−$12	.20	.20	.20	.20	.20
12	$5	$4	−$4	−$5	−$8	.20	.20	.20	.20	.20	$4	$2	−$2	−$4	−$8	.20	.20	.20	.20	.20
13	$15	$0	−$15			.20	.40	.40			$10	$0	−$10	−$15		.20	.40	.20	.20	
14	$10	$0	−$10	−$12		.20	.40	.20	.20		$5	$0	−$5	−$12		.20	.40	.20	.20	
15	$5	$0	−$5	−$6		.20	.40	.20	.20		$2	$0	−$2	−$6		.20	.40	.20	.20	
16	$15	$0	−$15			.40	.40	.20			$15	$10	$0	−$10		.20	.20	.40	.20	
17	$6	$5	$0	−$5		.20	.20	.40	.20		$6	$2	$0	−$2		.20	.20	.40	.20	
18	$15	$0	−$15			.20	.20	.60			$10	$0	−$10	−$15		.20	.20	.20	.40	
19	$10	$0	−$10	−$12		.20	.20	.20	.40		$5	$0	−$5	−$12		.20	.20	.20	.40	
20	$5	$0	−$5	−$6		.20	.20	.20	.40		$2	$0	−$2	−$6		.20	.20	.20	.40	
21	$15	$0	−$15			.60	.20	.20			$15	$10	$0	−$10		.40	.20	.20	.20	
22	$12	$10	$0	−$10		.40	.20	.20	.20		$12	$5	$0	−$5		.40	.20	.20	.20	
23	$6	$5	$0	−$5		.40	.20	.20	.20		$6	$2	$0	−$2		.40	.20	.20	.20	
24	$15	$0	−$15			.20	.60	.20			$10	$0	−$10			.20	.60	.20		
25	$10	$0	−$10			.20	.60	.20			$5	$0	−$5			.20	.60	.20		
26	$5	$0	−$5			.20	.60	.20			$2	$0	−$2			.20	.60	.20		
27	$15	$9	$0	−$9	−$15	.20	.20	.20	.20	.20	$10	$9	$0	−$9	−$10	.20	.20	.20	.20	.20
28	$10	$3	$0	−$3	−$10	.20	.20	.20	.20	.20	$5	$3	$0	−$3	−$5	.20	.20	.20	.20	.20
29	$5	$2	$0	−$2	−$5	.20	.20	.20	.20	.20	$2	$0	−$2			.40	.20	.40		
30	$15	$8	$5	$0	−$15	.20	.20	.20	.20	.20	$10	$8	$5		−$10	.20	.20	.20	.20	.20
31	$10	$4	$2	$0	−$10	.20	.20	.20	.20	.20	$5	$4	$2		−$5	.20	.20	.20	.20	.20
32	$5	$2	$1	$0	−$5	.20	.20	.20	.20	.20	$2	$1	$0	−$2		.40	.20	.20	.20	
33	$15	$0	−$5	−$8	−$15	.20	.20	.20	.20	.20	$10	$0	−$5	−$8	−$10	.20	.20	.20	.20	.20
34	$10	$0	−$2	−$4	−$10	.20	.20	.20	.20	.20	$5	$0	−$2	−$4	−$5	.20	.20	.20	.20	.20
35	$5	$0	−$1	−$2	−$5	.20	.20	.20	.20	.20	$2	$0	−$1	−$2		.20	.20	.20	.40	
36	$15	$5	−$15			.40	.20	.40			$15	$10	$5	−$10	−$15	.20	.20	.20	.20	.20
37	$10	$3	−$10			.40	.20	.40			$10	$5	$3	−$5	−$10	.20	.20	.20	.20	.20
38	$5	$2	−$5			.40	.20	.40			$5	$2	−$2	−$5		.20	.40	.20	.20	
39	$15	$0	−$15			.40	.20	.40			$15	$10	$0	−$10	−$15	.20	.20	.20	.20	.20
40	$10	$0	−$10			.40	.20	.40			$15	$10	$0	−$10	−$15	.20	.20	.20	.20	.20
41	$5	$0	−$5			.40	.20	.40			$5	$2	$0	−$2	−$5	.20	.20	.20	.20	.20
42	$15	−$5	−$15			.40	.20	.40			$15	$10	−$5	−$10	−$15	.20	.20	.20	.20	.20
43	$10	−$3	−$10			.40	.20	.40			$15	$10	−$3	−$10	−$15	.20	.20	.20	.20	.20
44	$5	−$2	−$5			.40	.20	.40			$5	$2	−$2	−$5		.20	.20	.40	.20	

A.4. True-and-error modeling

A rigorous method of evaluating the presence of paradoxical preference without making reference to models such as CPT, TAX, or DFT$_e$ is through the use of *true-and-error models* (TE models; for an overview, see Birnbaum, 2013). These models describe response patterns across a small set of lottery problems in terms of: (a) latent-preference states S, and (b) error probabilities ϵ. For example, consider the following four possible preference states pertaining to two lottery problems: $S_{\mathcal{A},\mathcal{A}'}$, $S_{\mathcal{B},\mathcal{B}'}$, $S_{\mathcal{A},\mathcal{B}'}$, and $S_{\mathcal{B},\mathcal{A}'}$. The first and second subscripts in S denote the preferences expressed for the lottery problems $\{\mathcal{A}, \mathcal{B}\}$ and $\{\mathcal{A}', \mathcal{B}'\}$. For example, the subscript $\mathcal{A}, \mathcal{B}'$ or response pattern $\mathcal{A}\mathcal{B}'$ are both shorthand for '$\mathcal{A} \succ \mathcal{B}$ *and* $\mathcal{A}' \prec \mathcal{B}'$' (the first at a latent level, the second at the level of observed choices). Note that when applied to the choice pairs from the set of 'new paradoxes,' states $S_{\mathcal{A},\mathcal{B}'}$ and $S_{\mathcal{B},\mathcal{A}'}$ are paradoxical (i.e., not allowed) under CPT. An individual follows in a given state with some probability $P(S)$, with $P(S_{\mathcal{A},\mathcal{A}'}) + P(S_{\mathcal{B},\mathcal{B}'}) + P(S_{\mathcal{A},\mathcal{B}'}) + P(S_{\mathcal{B},\mathcal{A}'}) = 1$. When individuals attempt to express their preferences, they can fail to do so with some probability $0 \leqslant \epsilon \leqslant .50$, a probability that can vary across lottery pairs. Because of the possibility of errors, the observed choice pattern across pairs can be expressed as a mixture of the different preference states. Importantly, TE models make no claims about the exact nature of these error probabilities. As discussed by Birnbaum (2013), "*A person might misread the problem, misremember the information, misaggregate the information, misremember the decision, or push the wrong key*" (p. 718).

As an example, consider the probability of observing response pattern $\mathcal{A}\mathcal{B}'$ when presented with the two lottery problems:

$$P(\mathcal{A}\mathcal{B}') = P(S_{\mathcal{A},\mathcal{A}'})\cdot(1-\epsilon_{\mathcal{A}\mathcal{B}})\cdot\epsilon_{\mathcal{A}'\mathcal{B}'} + P(S_{\mathcal{B},\mathcal{B}'})\cdot\epsilon_{\mathcal{A}\mathcal{B}}\cdot(1-\epsilon_{\mathcal{A}'\mathcal{B}'}) + P(S_{\mathcal{A},\mathcal{B}'})\cdot(1-\epsilon_{\mathcal{A}\mathcal{B}})\cdot(1-\epsilon_{\mathcal{A}'\mathcal{B}'}) + P(S_{\mathcal{B},\mathcal{A}'})\cdot\epsilon_{\mathcal{A}\mathcal{B}}\cdot\epsilon_{\mathcal{A}'\mathcal{B}'} \quad (15)$$

We can test for the presence of paradoxical preferences by comparing unconstrained true-and-error models against constrained

Table A3
Sets of Lottery Problems used to Test Choice Paradoxes (Gain and Mixed Domains).

	Lottery $\mathcal{A}/\mathcal{A}'/\mathcal{A}''$								Lottery $\mathcal{B}/\mathcal{B}'/\mathcal{B}''$							
Paradox	x_1	x_2	x_3	x_4	p_1	p_2	p_3	p_4	x_1	x_2	x_3	x_4	p_1	p_2	p_3	p_4
ES_{G1}	\$100	\$50	\$50		.85	.10	.05		\$100	\$100	\$8		.85	.10	.05	
	\$100	\$50			.85	.15			\$100	\$8			.95	.05		
ES_{G2}	\$98	\$2			.10	.90			\$40	\$2			.20	.80		
	\$98	\$2	\$2		.10	.10	.80		\$40	\$40	\$2		.10	.10	.80	
ES_{G3}	\$98	\$98	\$2		.80	.10	.10		\$98	\$40	\$40		.80	.10	.10	
	\$98	\$2			.90	.10			\$98	\$40			.80	.20		
ES_M	\$100	\$0	− \$50	− 50	.25	.25	.25	.25	\$50	\$50	\$0	− \$100	.25	.25	.25	.25
	\$100	\$0	− \$50		.25	.25	.50		\$50	\$0	− \$100		.50	.25	.25	
SD_G	\$96	\$14	\$12		.90	.05	.05		\$96	\$90	\$12		.85	.05	.10	
	\$97	\$15	\$13		.90	.05	.05		\$90	\$80	\$10		.85	.05	.10	
UTI_{G1}	\$92	\$68	\$0		.43	.07	.50		\$92	\$0			.48	.52		
	\$98	\$68	\$0		.43	.07	.50		\$98	\$92	\$0		.43	.05	.52	
UTI_{G2}	\$110	\$44	\$40		.80	.10	.10		\$110	\$96	\$10		.80	.10	.10	
	\$96	\$44	\$40		.80	.10	.10		\$96	\$10			.90	.10		
LCI_{G1}	\$52	\$48	\$4		.05	.05	.90		\$96	\$12	\$4		.05	.05	.90	
	\$52	\$12			.10	.90			96	\$12			.05	.95		
LCI_{G2}	\$44	\$40	\$2		.10	.10	.80		\$98	\$10	\$2		.10	.10	.80	
	\$44	\$10			.20	.80			\$98	\$10			.10	.90		
LCI_{M1}	\$19	\$15	− \$23		.10	.10	.80		\$73	− \$15	− \$23		.10	.10	.80	
	\$19	− \$15			.20	.80			\$73	− \$15			.10	.90		
LCI_{M2}	\$27	\$23	− \$22		.05	.05	.90		\$71	− \$13	− \$22		.05	.05	.90	
	\$27	− \$13			.10	.90			\$71	− \$13			.05	.95		
UCI_{G1}	\$110	\$44	\$40		.80	.10	.10		\$110	\$98	\$10		.80	.10	.10	
	\$98	\$40			.80	.20			\$98	\$10			.90	.10		
UCI_{G2}	\$106	\$52	\$48		.90	.05	.05		\$106	\$96	\$12		.90	.05	.05	
	\$96	\$48			.90	.10			\$96	\$12			.95	.05		
UCI_{L1}	\$60	− \$6	− \$10		.80	.10	.10		\$60	\$48	− \$40		.80	.10	.10	
	\$48	− \$10			.80	.20			\$48	− \$40			.90	.10		
UCI_{L2}	\$56	\$2	− \$2		.90	.05	.05		\$56	\$46	− \$38		.90	.05	.05	
	\$46	− \$2			.90	.10			\$46	− \$38			.95	.05		

Note. Each set corresponds to a pair of lottery problem, and their scaled replications (the latter are not reported here). The lottery problems used in the loss domain correspond to the ones used in the gain domain, but with all outcomes multiplied by − 1.

versions in which the probabilities of the paradoxical preference states are fixed to zero (i.e., $P(S_{\mathcal{A},\mathcal{B}'}) = P(S_{\mathcal{B},\mathcal{A}'}) = 0$). The testing of true-and-error models usually includes one replication of each lottery problems in a set, such that the pattern being tested consists of four choices (e.g. $\mathcal{AB}'\mathcal{AB}'$) rather than only two (e.g. $\mathcal{AB}'$), as in the example above.[19] It is commonly assumed that preferences do not change between replications (Birnbaum, 2013). The inclusion of replications along with this preference invariance assumption constraints the TE model and allows its testing. In the present work we followed a somewhat distinct approach in the sense that we used '*scaled replications*' in which all the outcomes in a given option pair are multiplied by a positive constant.[20] We allowed for error probabilities to differ across scaled replications. Overall, the unconstrained true-error model for a four-choice pattern consists of

[19] To facilitate comparisons, in the case of tests of stochastic dominance, we also considered a pair of lotteries problems and their respective scaled replications.

[20] If the utility function of gains/losses follows a power function ($u(x) = ax^b$), as assumed by all the models considered here, then preferences should remain invariant to such kind of scalings (for a proof, see Cho & Fisher, 2000). For example, the preferences for the pair $\mathcal{A}^* = \begin{pmatrix} \$78 & \$72 & \$4.5 \\ .05 & .05 & .90 \end{pmatrix}$ and $\mathcal{B}^* = \begin{pmatrix} \$144 & \$18 & \$4.5 \\ .05 & .05 & .90 \end{pmatrix}$ obtained by multiplying all outcomes of $\mathcal{A} = \begin{pmatrix} \$52 & \$48 & \$3 \\ .05 & .05 & .90 \end{pmatrix}$ and $\mathcal{B} = \begin{pmatrix} \$96 & \$12 & \$3 \\ .05 & .05 & .90 \end{pmatrix}$ by 1.5 should be the same as in the latter pair. Lopes and Oden (1999) report results in line with this scaling invariance assumption for moderate constants like the present one.

Table A4
Choice-pattern frequencies tested with true-and-error models.

Set	Choice Pattern															
	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
ES_{G1}	9	4	2	1	2	**15**	1	5	0	1	1	3	1	3	0	7
ES_{G2}	3	1	1	1	0	0	1	4	3	2	**7**	6	1	2	2	21
ES_{G3}	8	3	5	1	1	0	1	1	4	1	**13**	6	2	0	2	7
ES_{L1}	3	0	4	0	1	1	2	1	1	1	**14**	6	2	2	5	12
ES_{L2}	30	6	2	1	4	2	0	2	1	0	0	0	0	3	2	2
ES_{L3}	8	7	1	1	8	**15**	0	4	1	2	0	0	0	2	0	6
ES_{M}	27	3	2	2	6	3	3	1	1	0	1	2	1	2	0	1
SD_{G}	**17**	5	7	4	5	4	0	1	2	0	2	2	1	2	0	3
SD_{L}	2	0	0	1	3	3	1	1	0	3	7	5	3	1	10	**15**
UTI_{G1}	5	2	4	2	3	1	3	1	2	1	**11**	5	0	0	6	9
UTI_{G2}	5	6	0	2	9	**13**	2	2	0	1	0	3	2	6	0	4
UTI_{L1}	9	3	1	0	4	**9**	2	1	1	2	1	1	4	3	5	9
UTI_{L2}	1	4	2	0	2	0	2	0	3	2	**13**	4	0	0	7	15
LCI_{G1}	10	4	6	2	3	3	0	1	3	2	6	2	2	5	2	4
LCI_{G2}	29	0	2	3	0	1	1	1	4	0	**10**	2	0	0	1	1
LCI_{L1}	2	3	4	2	1	5	1	4	2	1	1	3	1	3	6	16
LCI_{L2}	0	2	0	2	1	1	2	6	1	1	0	3	3	4	2	27
LCI_{M1}	15	1	5	0	3	0	1	1	4	0	**6**	5	1	1	3	9
LCI_{M2}	11	1	2	3	1	1	1	2	3	2	**11**	7	1	0	4	5
UCI_{G1}	5	8	1	2	7	**14**	1	3	1	0	1	2	2	4	0	4
UCI_{G2}	4	6	2	1	5	**9**	3	7	2	1	1	0	0	7	1	6
UCI_{L1}	4	0	0	0	1	2	2	0	2	0	**19**	5	0	1	8	11
UCI_{L2}	6	0	3	2	1	0	3	2	3	2	**14**	4	1	0	5	9
UCI_{M1}	0	11	6	18	0	4	1	6	0	2	0	2	0	1	0	4
UCI_{M2}	4	0	4	0	6	0	5	0	2	0	2	0	15	1	15	1

Note. Response patterns for each set testing a given choice paradox in a given domain (see Table A3). Pattern digits 1/0 represent a choice for $\mathcal{A}/\mathcal{B}$ in a given lottery problem. The third and fourth digits correspond to the scaled replications of the first and second lottery problems. ESC = Event splitting and coalescing, SD = First-order stochastic dominance, UTI = upper tail independence, LCI = lower-cumulative independence, UCI = upper cumulative independence.. Subscripts G, L, and M denote gain, loss, and mixed domains, respectively. The frequencies in bold highlight frequencies in which paradoxical choice patterns were perfectly replicated. Note that the paradoxical patterns in the loss domain mirror the gain domain.

seven free parameters, three defining the four preference-state probabilities (again, note that they sum to 1), and four parameter capturing the error probabilities for each lottery problem. For example, the unconstrained TE model describes the probability of response pattern as follows (with $\star$ denoting the scaled replications):

$$P(\mathcal{A}\mathcal{B}'\mathcal{A}^{\star}\mathcal{B}^{\star'}) = P(S_{\mathcal{A},\mathcal{A}'})\cdot(1-\epsilon_{\mathcal{A}\mathcal{B}})\cdot\epsilon_{\mathcal{A}'\mathcal{B}'}\cdot(1-\epsilon^{\star}_{\mathcal{A}\mathcal{B}})\cdot\epsilon^{\star}_{\mathcal{A}'\mathcal{B}'} + P(S_{\mathcal{B},\mathcal{B}'})\cdot\epsilon_{\mathcal{A}\mathcal{B}}\cdot(1-\epsilon_{\mathcal{A}'\mathcal{B}'})\cdot\epsilon^{\star}_{\mathcal{A}\mathcal{B}}\cdot(1-\epsilon^{\star}_{\mathcal{A}'\mathcal{B}'})$$
$$+ P(S_{\mathcal{A},\mathcal{B}'})\cdot(1-\epsilon_{\mathcal{A}\mathcal{B}})\cdot(1-\epsilon_{\mathcal{A}'\mathcal{B}'})\cdot(1-\epsilon^{\star}_{\mathcal{A}\mathcal{B}})\cdot(1-\epsilon^{\star}_{\mathcal{A}'\mathcal{B}'}) + P(S_{\mathcal{B},\mathcal{A}'})\cdot\epsilon_{\mathcal{A}\mathcal{B}}\cdot\epsilon_{\mathcal{A}'\mathcal{B}'}\cdot\epsilon^{\star}_{\mathcal{A}\mathcal{B}}\cdot\epsilon^{\star}_{\mathcal{A}'\mathcal{B}'}$$

The sets of lottery problems used in the tests are provided in Table A3. In order to test the choice paradoxes, we compared the performance (using Bayes factors) of three models: the constrained and unconstrained true-and-error models, and a saturated model that can accommodate any choice pattern. Different constrained models were considered, depending on the type of paradox. In all cases, the constraint consisted of setting the probabilities of preference states to 0. A superior performance by the constrained model indicates that there is no need for assuming paradoxical preference states, whereas a superior performance by the unconstrained model would make the opposite case. A victorious saturated model would indicate that none of the true-and-error models can successfully capture the choice data.

Table A4 reports the choice patterns fitted, and Table A5 reports the parameter estimates of the unconstrained model, as well the log Bayes Factors comparing the different models. By and large, the results supported the notion that *many* individuals hold preferences that are paradoxical under CPT, with paradoxical preferences often being attributed large probabilities. These results reflect the fact that many participants not only produced 'paradoxical' choice patterns, but repeated them exactly in a scaled replication. Notably, the saturated model was preferred when testing violations of upper-cumulative independence in the domain of mixed lotteries. As can be seen from the choice patterns reported in Table A4, the failure of the unconstrained model can be attributed to the fact that many participants showed a clear preference reversal across scaled replicates.

Table A5
Parameter estimates of the unconstrained TE model and model-selection results.

Set	$P(S_{\mathcal{A},\mathcal{A}'})$	$P(S_{\mathcal{A},\mathcal{B}'})$	$P(S_{\mathcal{B},\mathcal{A}'})$	$P(S_{\mathcal{B},\mathcal{B}'})$	$\epsilon_{\mathcal{AB}}$	$\epsilon_{\mathcal{A}'\mathcal{B}'}$	$\epsilon^{\star}_{\mathcal{AB}}$	$\epsilon^{\star}_{\mathcal{A}'\mathcal{B}'}$	$\log\text{BF}_{u,c}$	$\log\text{BF}_{u,s}$
ES_{G1}	.24 [.11,.38]	**.04 [.00,.13]**	**.47 [.31,.63]**	.25 [.13,.39]	.11 [.01,.27]	.22 [.08,.35]	.22 [.07,.36]	.09 [.00,.24]	11.47	6.43
ES_{G2}	.65 [.49,.81]	**.21 [.06,.37]**	**.03 [.00,.10]**	.11 [.03,.23]	.17 [.07,.30]	.22 [.06,.38]	.15 [.05,.29]	.12 [.01,.30]	1.84	3.44
ES_{G3}	.19 [.06,.34]	**.48 [.31,.68]**	**.03 [.00,.11]**	.29 [.15,.45]	.19 [.03,.37]	.13 [.01,.30]	.16 [.01,.34]	.26 [.11,.41]	9.41	5.23
ES_{L1}	.37 [.18,.55]	**.48 [.30,.69]**	**.05 [.00,.14]**	.10 [.01,.22]	.16 [.05,.29]	.24 [.03,.45]	.14 [.04,.28]	.19 [.01,.43]	4.79	5.16
ES_{L2}	.14 [.05,.24]	**.02 [.00,.07]**	**.08 [.00,.21]**	.77 [.63,.89]	.07 [.01,.17]	.11 [.01,.23]	.13 [.05,.24]	.20 [.10,.33]	−1.48	−0.59
ES_{L3}	.14 [.05,.27]	**.02 [.00,.09]**	**.56 [.37,.78]**	.27 [.08,.44]	.12 [.02,.25]	.25 [.05,.43]	.14 [.03,.27]	.18 [.01,.39]	8.19	3.12
ES_{M}	.10 [.02,.20]	**.03 [.00,.12]**	**.08 [.00,.21]**	.79 [.64,.92]	.08 [.01,.19]	.27 [.14,.41]	.19 [.08,.31]	.15 [.02,.30]	−1.25	1.26
SD_{G}	.12 [.03,.25]	.06 [.00,.18]	.11 [.00,.27]	**.71 [.52,.87]**	.10 [.01,.24]	.18 [.03,.35]	.29 [.17,.43]	.26 [.11,.42]	29.28	4.41
SD_{L}	**.63 [.37,.86]**	.17 [.01,.42]	.12 [.01,.25]	.08 [.00,.21]	.08 [.00,.20]	.27 [.03,.47]	.16 [.06,.28]	.40 [.20,.50]	30.12	3.52
UTI_{G1}	.28 [.05,.51]	**.46 [.23,.73]**	**.06 [.00,.18]**	.20 [.07,.35]	.22 [.09,.36]	.31 [.07,.48]	.08 [.00,.22]	.27 [.04,.47]	3.77	6.67
UTI_{G2}	.13 [.01,.29]	**.05 [.00,.16]**	**.70 [.45, .91]**	.12 [.00, .34]	.22 [.09, .37]	.28 [.09, .46]	.14 [.02, .30]	.31 [.09, .48]	7.80	3.84
UTI_{L1}	.35 [.19, .52]	**.05 [.00, .15]**	**.33 [.15, .53]**	.28 [.11, .45]	.20 [.03, .36]	.14 [.01, .38]	.13 [.01, .32]	.36 [.20, .49]	3.70	4.68
UTI_{L2}	.38 [.11, .66]	**.43 [.16, .72]**	**.06 [.00, .18]**	.13 [.01, .26]	.10 [.01, .23]	.27 [.03, .48]	.11 [.02, .24]	.31 [.06, .49]	2.54	−0.47
LCI_{G1}	.21 [.03, .41]	**.20 [.01, .44]**	.14 [.01, .35]	.45 [.22, .69]	.25 [.02, .48]	.17 [.01, .38]	.33 [.06, .49]	.25 [.07, .42]	1.83	7.86
LCI_{G2}	.06 [.01, .14]	**.27 [.14, .41]**	.04 [.00, .11]	.63 [.49, .77]	.12 [.01, .27]	.05 [.00, .14]	.14 [.02, .29]	.12 [.04, .22]	8.05	−2.27
LCI_{L1}	.61 [.41, .81]	.09 [.00, .26]	**.14 [.01, .32]**	.16 [.03, .33]	.25 [.05, .44]	.19 [.03, .37]	.22 [.03, .41]	.25 [.09, .42]	0.45	5.78
LCI_{L2}	.88 [.74, .98]	.02 [.00, .09]	**.04 [.00, .17]**	.05 [.00, .15]	.24 [.12, .38]	.16 [.07, .28]	.20 [.08, .34]	.17 [.07, .29]	−2.10	0.91
LCI_{M1}	.29 [.16, .43]	**.23 [.09, .39]**	.03 [.00, .09]	.46 [.31, .61]	.14 [.01, .30]	.22 [.09, .35]	.16 [.02, .31]	.09 [.01, .22]	6.93	6.47
LCI_{M2}	.20 [.04, .37]	**.45 [.26, .66]**	.05 [.00, .14]	.31 [.16, .46]	.20 [.03, .36]	.17 [.01, .37]	.13 [.01, .31]	.30 [.12, .46]	8.54	4.77
UCI_{G1}	.12 [.01, .27]	.06 [.00, .17]	**.66 [.42, .89]**	.16 [.01, .37]	.17 [.04, .32]	.31 [.11, .47]	.17 [.04, .32]	.26 [.04, .45]	9.11	6.04
UCI_{G2}	.17 [.01, .39]	.06 [.00, .18]	**.59 [.33, .86]**	.18 [.01, .36]	.26 [.04, .45]	.19 [.02, .37]	.33 [.12, .48]	.23 [.06, .40]	6.52	5.46
UCI_{L1}	.29 [.12, .45]	**.55 [.38, .73]**	.06 [.01, .15]	.11 [.03, .21]	.06 [.00, .14]	.25 [.06, .43]	.08 [.01, .17]	.15 [.01, .37]	7.60	3.36
UCI_{L2}	.28 [.10, .45]	**.51 [.32, .74]**	.03 [.00, .10]	.18 [.05, .32]	.24 [.10, .39]	.25 [.05, .43]	.12 [.01, .28]	.19 [.01, .40]	8.16	4.87
UCI_{M1}	.09 [.00, .23]	.02 [.00, .09]	**.81 [.63, .97]**	.07 [.00, .19]	.11 [.01, .26]	.48 [.42, .50]	.47 [.42, .50]	.08 [.00, .21]	8.38	−8.11
UCI_{M2}	.25 [.00, .83]	.50 [.00, .97]	**.05 [.00, .24]**	.20 [.00, .57]	.26 [.03, .48]	.35 [.01, .50]	.42 [.24, .50]	.18 [.00, .50]	−1.89	−4.59

Note. Posterior parameter means (95% credibility intervals in squared brackets) for the unconstrained TE model. Values in bold indicate the preference states that are paradoxical in a given set (note that these can differ between gain and loss domains). The key for the paradox acronyms is given in Table A4. Subscripts *G*, *L*, and *M* denote gain, loss, and mixed domains, respectively. The log Bayes Factors compare the unconstrained TE model (subscript u), the constrained model (c), and a saturated model (s). Positive values indicate a support for the model denoted by the first subscript, whereas negative values indicate support for the model in the second subscript.

References

Abdellaoui, M., Bleichrodt, H., & L'Haridon, O. (2008). A tractable method to measure utility and loss aversion under prospect theory. *Journal of Risk and Uncertainty, 36*, 245–266.
Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école américaine. *Econometrica, 21*, 503–546.
Andraszewicz, S., Rieskamp, J., & Scheibehenne, B. (2015). How outcome dependencies affect decisions under risk. *Decision, 2*, 127–144.
Baucells, M., & Heukamp, F. H. (2006). Stochastic dominance and cumulative prospect theory. *Management Science, 52*, 1409–1423.
Bernoulli, D. (1738/1954). Exposition of a new theory on the measurement of risk. *Econometrica, 22*, 23–26.
Bhatia, S. (2014). Sequential sampling and paradoxes of risky choice. *Psychonomic Bulletin & Review, 21*, 1095–1111.
Bhatia, S., & Loomes, G. (2017). Noisy preferences in risky choice: A cautionary note. *Psychological Review, 124*, 678–687.
Birnbaum, M. H. (1973). The devil rides again: Correlation as an index of fit. *Psychological Bulletin, 79*, 239–242.
Birnbaum, M. H. (1997). Violations of monotonicity in judgment and decision making. In A. A. J. Marley (Ed.). *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 73–100). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
Birnbaum, M. H. (2005). Three new tests of independence that differentiate models of risky decision making. *Management Science, 51*, 1346–1358.
Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychological Review, 115*, 463–501.
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., & Chavez, A. (1997). Tests of theories of decision making: Violations of branch independence and distribution independence. *Organizational Behavior and Human Decision Processes, 71*, 161–194.
Birnbaum, M. H., & Jou, J.-W. (1990). A theory of comparative response times and difference judgments. *Cognitive Psychology, 22*, 184–210.
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*, 49–79.
Birnbaum, M. H., Parducci, A., & Gifford, R. K. (1971). Contextual effects in information integration. *Journal of Experimental Psychology, 88*, 158.
Birnbaum, M. H. (2004). Causes of Allais common consequence paradoxes: An experimental dissection. *Journal of Mathematical Psychology, 48*, 87–106.
Birnbaum, M. H., & Schmidt, U. (2010). Testing transitivity in choice under risk. *Theory and Decision, 69*, 599–614.
Birnbaum, M. H., & Schmidt, U. (2015). The impact of learning by thought on violations of independence and coalescing. *Decision Analysis, 12*, 144–152.
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.
Birnbaum, M. H., & Sutton, S. E. (1992). Scale convergence and utility measurement. *Organizational Behavior and Human Decision Processes, 52*, 183–215.
Blavatskyy, P. R. (2008). Stochastic utility theorem. *Journal of Mathematical Economics, 44*, 1049–1056.
Booij, A. S., & Van de Kuilen, G. (2009). A parameter-free analysis of the utility of money for the general population under prospect theory. *Journal of Economic Psychology, 30*, 651–666.
Brooks, P., Peters, S., & Zank, H. (2014). Risk behavior for gain, loss, and mixed prospects. *Theory and Decision, 77*, 153–182.
Broomell, S. B., & Bhatia, S. (2014). Parameter recovery for decision modeling using choice data. *Decision, 1*, 252–274.
Busemeyer, J. R., & Diederich, A. (2002). Survey of decision field theory. *Mathematical Social Sciences, 43*, 345–370.

Busemeyer, J. R., & Townsend, J. T. (1992). Fundamental derivations from decision field theory. *Mathematical Social Sciences, 23*, 255–282.
Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review, 100*, 432.
Cho, Y.-H., & Fisher, G. R. (2000). Receiving two consequences: Tests of monotonicity and scale invariance. *Organizational Behavior and Human Decision Processes, 83*, 61–81.
Costello, F., & Watts, P. (2014). Surprisingly rational: Probability theory plus noise explains biases in judgment. *Psychological Review, 121*, 463–480.
Costello, F., & Watts, P. (2016). People's conditional probability judgments follow probability theory (plus noise). *Cognitive Psychology, 89*, 106–133.
Davis-Stober, C. P., & Brown, N. (2013). Evaluating decision maker "type" under p-additive utility representations. *Journal of Mathematical Psychology, 57*(6), 320–328.
Diederich, A. (2003a). Decision making under conflict: Decision time as a measure of conflict strength. *Psychonomic Bulletin & Review, 10*, 167–176.
Diederich, A. (2003b). Mdft account of decision making under time pressure. *Psychonomic Bulletin & Review, 10*, 157–166.
Dougherty, M. R. P., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review, 106*, 180–209.
Dror, I. E., Basola, B., & Busemeyer, J. R. (1999). Decision making under time pressure: An independent test of sequential sampling models. *Memory & Cognition, 27*, 713–725.
Edwards, W. (1954). The theory of decision making. *Psychological Bulletin, 51*, 380.
Edwards, W. (1962). Subjective probabilities inferred from decisions. *Psychological Review, 69*, 109.
Erev, I., Ert, E., Plonsky, O., Cohen, D., & Cohen, O. (2017). From anomalies to forecasts: Toward a descriptive model of decisions under risk, under ambiguity, and from experience. *Psychological Review, 124*, 369–409.
Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over-and underconfidence: The role of error in judgment processes. *Psychological Review, 101*, 519–527.
Ert, E., & Erev, I. (2013). On the descriptive value of loss aversion in decisions under risk: Six clarifications. *Judgment and Decision Making, 8*, 214–235.
Fehr-Duda, H., & Epper, T. (2012). Probability and risk: Foundations and economic implications of probability-dependent risk preferences. *Annual Review of Economics, 4*, 567–593.
Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis.* Boca Raton, FL: CRC Press Boca Raton, FL.
Glöckner, A., Hilbig, B. E., Henninger, F., & Fiedler, S. (2016). The reversed description-experience gap: Disentangling sources of presentation format effects in risky choice. *Journal of Experimental Psychology: General, 145*, 486–508.
Glöckner, A., & Pachur, T. (2012). Cognitive models of risky choice: Parameter stability and predictive accuracy of prospect theory. *Cognition, 123*, 21–32.
Goldstein, W. M., & Einhorn, H. J. (1987). Expression theory and the preference reversal phenomena. *Psychological Review, 94*, 236–254.
Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology, 38*, 129–166.
Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science, 27*, 591–635.
Green, J. R., & Jullien, B. (1988). Ordinal independence in nonlinear utility theory. *Journal of Risk and Uncertainty, 1*, 355–387.
Griffiths, T. L., Vul, E., & Sanborn, A. N. (2012). Bridging levels of analysis for probabilistic models of cognition. *Current Directions in Psychological Science, 21*, 263–268.
Gronau, Q. F. & Singmann, H. (2017). Bridgesampling: Bridge sampling for marginal likelihoods and Bayes factors. R package version 0.4-0. Retrieved from <https://CRAN.R-project.org/package=bridgesampling>.
Hamilton, D. L., & Zanna, M. P. (1972). Differential weighting of favorable and unfavorable attributes in impressions of personality. *Journal of Experimental Research in Personality, 6*, 204–212.
Hartig, F., Minunno, F., Paul, S., Cameron, D., & Ott, T. (2017). Bayesiantools: General-purpose MCMC and SMC samplers and tools for Bayesian statistics. R package version 0.4-1. Retrieved from <https://CRAN.R-project.org/package=BayesianTools>.
Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science, 15*, 534–539.
Hilbert, M. (2012). Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making. *Psychological Bulletin, 138*, 211.
Hotaling, J. M., & Busemeyer, J. R. (2012). DFT-D: A cognitive-dynamical model of dynamic decision making. *Synthese, 189*, 67–80.
Humphrey, S. J. (1995). Regret aversion or event-splitting effects? More evidence under risk and uncertainty. *Journal of Risk and Uncertainty, 11*, 263–274.
Humphrey, S. J. (2001). Non-transitive choice: Event-splitting effects or framing effects? *Economica, 22*, 77–96.
Ingersoll, J. (2008). Non-monotonicity of the Tversky-Kahneman probability-weighting function: A cautionary note. *European Financial Management, 14*, 385–390.
Johnson, J. G., & Busemeyer, J. R. (2005). A dynamic, stochastic, computational model of preference reversal phenomena. *Psychological Review, 112*, 841–861.
Johnson, J. G., & Busemeyer, J. R. (2016). A computational model of the attention process in risky choice. *Decision, 3*, 254–280.
Kahneman, D. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.). *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). New York: Cambridge University Press.
Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263–291.
Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*, 773–795.
Kellen, D., Klauer, K. C., & Bröder, A. (2013). Recognition memory models and binary-response ROCs: A comparison by minimum description length. *Psychonomic Bulletin & Review, 20*, 693–719.
Kellen, D., Mata, R., & Davis-Stober, C. P. (2017). Individual classification of strong risk attitudes: An application across lottery types and age groups. *Psychonomic Bulletin & Review, 24*, 1341–1349.
Kellen, D., Pachur, T., & Hertwig, R. (2016). How (in) variant are subjective representations of described and experienced risk and rewards? *Cognition, 157*, 126–138.
Köbberling, V., & Wakker, P. P. (2005). An index of loss aversion. *Journal of Economic Theory, 122*, 119–131.
Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course.* Cambridge University Press.
Lerner, J. S., Small, D. A., & Loewenstein, G. (2004). Heart strings and purse strings: Carryover effects of emotions on economic decisions. *Psychological Science, 15*, 337–341.
Lieder, F., Griffiths, T. L., & Hsu, M. (2017). Overrepresentation of extreme events in decision making reflects rational use of cognitive resources. *Psychological Review*.
Lin, D., Donkin, C., & Newell, B. R. (2015). The exemplar confusion model: An account of biased probability estimates in decisions from description. In D. Noelle,, R. Dale,, A. Warlaumont,, J. Yoshimi,, T. Matlock,, & C. Jennings, (Eds.). *Proceedings of the 37th annual conference of the cognitive science society* (pp. 807–812). Mahwah, NJ: Erlbaum.
Loewenstein, G., O'Donoghue, T., & Bhatia, S. (2015). Modeling the interplay between affect and deliberation. *Decision, 2*(2), 55–81.
Loomes, G. (2005). Modelling the stochastic component of behaviour in experiments: Some issues for the interpretation of data. *Experimental Economics, 8*, 301–323.
Loomes, G., Moffatt, P. G., & Sugden, R. (2002). A microeconometric test of alternative stochastic theories of risky choice. *Journal of Risk and Uncertainty, 24*, 103–130.
Lopes, L. L., & Oden, G. C. (1999). The role of aspiration level in risky choice: A comparison of cumulative prospect theory and SP/A theory. *Journal of Mathematical Psychology, 43*, 286–313.
Luce, R. D. (1953). *Individual choice behavior: A theoretical analysis.* London: Dover.
Luce, R. D. (2000). *Utility of gains and losses: Measurement-theoretical and experimental approaches.* New York: Psychology Press.
Luce, R. D., & Marley, A. A. (2005). Ranked additive utility representations of gambles: Old and new axiomatizations. *Journal of Risk and Uncertainty, 30*, 21–62.
Marchiori, D., Di Guida, S., & Erev, I. (2015). Noisy retrieval models of over-and undersensitivity to rare events. *Decision, 2*, 82–106.
Marley, A. A., & Luce, R. D. (2005). Independence properties vis-à-vis several utility representations. *Theory and Decision, 58*(1), 77–143.
Meng, X.-L., & Schilling, S. (2002). Warp bridge sampling. *Journal of Computational and Graphical Statistics, 11*, 552–586.
Meng, X.-L., & Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica,* 831–860.
Millroth, P., Nilsson, H., & Juslin, P. (2018). Examining the integrity of evaluations of risky prospects using a single-stimuli design. *Decision, 5*, 362–377.
Mukherjee, K. (2010). A dual system model of preferences under risk. *Psychological Review, 117*, 243–255.
Myung, J. I., Cavagnaro, D. R., & Pitt, M. A. (2017). Model evaluation and selection. In W. H. Batchelder, H. Colonius, E. Dzhafarov, & J. I. Myung (Vol. Eds.), *Measurement and Methodology: . vol. 1. New Handbook of Mathematical Psychology* (pp. 552–598). Cambridge, UK: Cambridge University Press.
Nilsson, H., Rieskamp, J., & Wagenmakers, E.-J. (2011). Hierarchical Bayesian parameter estimation for cumulative prospect theory. *Journal of Mathematical Psychology, 55*, 84–93.
Pachur, T., Mata, R., & Hertwig, R. (2017). Who dares, who errs? Disentangling cognitive and motivational roots of age differences in decisions under risk.

Psychological Science, 28, 504–518.
Patalano, A. L., Saltiel, J. R., Machlin, L., & Barth, H. (2015). The role of numeracy and approximate number system acuity in predicting value and probability distortion. *Psychonomic Bulletin & Review, 22*, 1820–1829.
Payne, J. W. (2005). It is whether you win or lose: The importance of the overall probabilities of winning or losing in risky choice. *Journal of Risk and Uncertainty, 30*, 5–19.
Payne, J. W., Laughhunn, D. J., & Crum, R. (1980). Translation of gambles and aspiration level effects in risky choice behavior. *Management Science, 26*, 1039–1060.
Peirce, J. W. (2007). Psychopy—psychophysics software in python. *Journal of Neuroscience Methods, 162*, 8–13.
Platt, J. R. (1964). Strong inference. *Science, 146*, 347–353.
Prelec, D. (1998). The probability weighting function. *Econometrica, 66*, 497–527.
Pratto, F., & John, O. P. (1991). Automatic vigilance: the attention-grabbing power of negative social information. *Journal of Personality and Social Psychology, 61*, 380–391.
Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization, 3*, 323–343.
Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review, 111*, 333.
Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2011). Transitivity of preferences. *Psychological Review, 118*, 42–56.
Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1446–1465.
Rotello, C. M., Heit, E., & Dubé, C. (2015). When more data steer us wrong: replications with the wrong dependent measure perpetuate erroneous conclusions. *Psychonomic Bulletin & Review, 22*, 944–954.
Scheibehenne, B., & Pachur, T. (2015). Using bayesian hierarchical parameter estimation to assess the generalizability of cognitive models of choice. *Psychonomic Bulletin & Review, 22*, 391–407.
Scheibehenne, B., Rieskamp, J., & González-Vallejo, C. (2009). Cognitive models of choice: comparing decision field theory to the proportional difference model. *Cognitive Science, 33*, 911–939.
Schley, D. R., & Peters, E. (2014). Assessing economic value: Symbolic-number mappings predict risky and riskless valuations. *Psychological Science, 25*, 753–761.
Spektor, M. S., Kellen, D., & Hotaling, J. M. (2018). When the good looks bad: An experimental exploration of the repulsion effect. *Psychological science, 29*(8), 1309–1320.
Starmer, C., & Sugden, R. (1993). Testing for juxtaposition and event-splitting effects. *Journal of Risk and Uncertainty, 6*, 235–254.
Starns, J. J., Ratcliff, R., & McKoon, G. (2012). Evaluating the unequal-variance and dual-process explanations of zroc slopes with response time data and the diffusion model. *Cognitive Psychology, 64*, 1–34.
Stewart, N., Reimers, S., & Harris, A. J. (2014). On the origin of utility, weighting, and discounting functions: how they get their shapes and how to change their shapes. *Management Science, 61*, 687–705.
Stewart, N., & Simpson, K. (2008). A decision-by-sampling account of decision under risk. In N. Chater, & M. Oaksford (Eds.). *A decision-by-sampling account of decision under risk* (pp. 261–276). Oxford, UK: Oxford University Press.
ter Braak, C. J., & Vrugt, J. A. (2008). Differential evolution markov chain with snooker updater and fewer chains. *Statistics and Computing, 18*, 435–446.
Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*, 297–323.
Viscusi, W. K. (1989). Prospective reference theory: Toward an explanation of the paradoxes. *Journal of Risk and Uncertainty, 2*, 235–263.
von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior.* Princeton, NJ: Princeton University Press.
Wakker, P. P. (2010). *Prospect theory: For risk and ambiguity.* Cambridge, Mass.: Cambridge University Press.
Walasek, L., & Stewart, N. (2015). How to make loss aversion disappear and reverse: Tests of the decision by sampling origin of loss aversion. *Journal of Experimental Psychology: General, 144*, 7–11.
Wang, X. T., & Johnson, J. G. (2012). A tri-reference point theory of decision making under risk. *Journal of Experimental Psychology: General, 141*, 743.
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.
Zeisberger, S., Vrecko, D., & Langer, T. (2012). Measuring the time stability of prospect theory preferences. *Theory and Decision, 72*, 359–386.