Theoretical Note: An Ontology of Decision Models

Lisheng He, Wenjia Joyce Zhao, Sudeep Bhatia

University of Pennsylvania

Author Note

The ideas and part of the data in the manuscript were presented in Annual Meeting of Neuroeconomics Society (2018), Annual Meeting of Psychonomic Society (2018), and Annual Meeting of Cognitive Science Society (2019). Funding for the research was received from the National Science Foundation grant SES-1847794.

Correspondence concerning this manuscript should be addressed to: Lisheng He, Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104, USA. Email: hlisheng@sas.upenn.edu.

Abstract

Decision models are essential theoretical tools in the study of choice behavior, but there is little consensus about the best model for describing choice, with different fields and different research programs favoring their own idiosyncratic sets of models. Even within a given field, decision models are seldom studied alongside each other, and insights obtained using one model are not typically generalized to others. We present the results of a large-scale computational analysis that uses *landscaping* techniques to generate a representational structure for describing decision models. Our analysis includes 89 prominent models of risky and intertemporal choice, and results in an ontology of decision models, interpretable in terms of model spaces, clusters, hierarchies, and graphs. We use this ontology to measure the properties of individual models and quantify the relationships between different models. Our results show how decades of quantitative research on human choice behavior can be synthesized within a single representational framework.

Keywords: decision making, risky choice, intertemporal choice, meta-theoretical analysis, modeling

Introduction

The study of how people make decisions is a central topic of research in psychology, as well as in various other social, behavioral, and biological sciences (Kahneman & Tversky, 2000; Camerer, Loewenstein, & Rabin, 2011; Glimcher & Fehr, 2013). This research has been remarkably influential, shaping our understanding of the psychological determinants of choice, of individual rationality, of markets and societies, and of the biological bases of human behavior (Bettman, Luce, & Payne, 1998; Starmer, 2000; Glimcher & Rustichini, 2004; Weber & Johnson, 2009). Much of this work has relied on *decision models* in order to describe choice processes, predict choice outcomes, and interpret the relationship between choices and various affective, cognitive, clinical, socioeconomic, demographic, and neurobiological variables.

Decision models are parameterized mathematical functions or computer algorithms, which take as inputs a set of available choice options and produce as outputs predictions regarding decision makers' choices over this set. In risky decision making, for example, these models predict choices over gambles, which offer potentially probabilistic rewards. Likewise, in intertemporal decision making, these models predict choices over sequences of outcomes, which offer potentially delayed rewards. By quantitatively describing the ways in which choices are made, decision models allow researchers to infer parameters corresponding to latent decision constructs (risk aversion, time discounting, regret, probability weighting, attentional bias, loss aversion, present bias etc.) from behavioral data, giving the study of decision making conceptual rigor and empirical precision. For this reason, decision models are essential theoretical tools in psychology (Birnbaum, 2008; Busemeyer & Townsend, 1993; Brandstätter, Gigerenzer, & Hertwig, 2006; Scholten & Read, 2010; Ericson, White, Laibson, & Cohen, 2015), economics (Loomes & Sugden, 1982; Tversky & Kahneman, 1992; Loewenstein & Prelec, 1992; Laibson, 1997; Yaari, 1987) and neuroscience (McClure, Ericson, Laibson, Loewenstein, & Cohen, 2007; Kable & Glimcher, 2010), and have been extensively applied to study human behavior in clinical, financial, managerial, consumer, policy, and other applied domains.

However, despite decades of decision research, we do not currently have a unified model of choice behavior, or any academic consensus about the right decision model for studying how people make decisions. Rather, the long history and vast interdisciplinary scope of decision research have given rise to a very large number of distinct models, each making seemingly unique claims about how people deliberate and choose between available options. In this paper, we catalogue 89 different models of simple risky and intertemporal choice. The existence of so many decision models complicates our understanding of choice behavior, impeding scientific progress.

Another source of confusion is the fact that decision models involve a menagerie of overlapping assumptions. In risky choice, for example, some decision models may assume a non-linear transformation of payoffs (e.g. expected utility theory, Bernoulli, 1738), others may assume a non-linear transformation of probabilities (e.g. dual theory, Yaari, 1987), and some may assume both (e.g. cumulative prospect theory, Tversky & Kahneman, 1992). Likewise, some models may allow the payoffs offered by a gamble to influence how other payoffs of the same gamble are evaluated (e.g. the transfer-of-attention exchange model, Birnbaum, 2008), some may allow the payoffs of a gamble to influence how payoffs of other gambles are evaluated (e.g. regret theory, Loomes & Sugden, 1982); and others may do both (e.g. the priority heuristic, Brandstatter et al., 2006). Which of these assumptions give rise to the idiosyncratic predictions of the model, and do two models that share a given assumption make similar predictions?

Without answering these questions, our understanding of the essential mathematical operations necessary to describe human choice behavior remains incomplete.

Finally, it is difficult to make rigorous model-based empirical claims without first characterizing the relationships between different decision models and between the constructs that their parameters represent. Imagine observing a relationship between a model parameter and an affective, cognitive, clinical, socioeconomic, demographic, or neurobiological variable of interest. For example, activity in the limbic system may correlate with the decision maker's weighting of immediate payoffs (McClure, Laibson, Loewenstein, & Cohen, 2004), time pressure may be associated with the use of a particular heuristic (Payne, Bettman, & Johnson, 1988), higher incentives may lead to an increase in risk aversion (Holt & Laury, 2002), and individuals prone to addictive behavior may be more likely to discount future rewards (MacKillop et al., 2011). Testing for such relationships is increasingly common and these tests represent one of the main ways in which decision models are used to describe empirical regularities in choice behavior. However, these tests typically involve the parameters or predictions of a limited number of models or even a single model, and we cannot tell if the variable of interest is better described by one of the numerous other models in the literature. A range of different decision constructs (specified by a range of different models) could be associated with limbic system activation, the effects of time pressure and incentives, and addiction proneness. Understanding these associations is necessary for a rigorous, cumulative, and trans-disciplinary science of human choice behavior.

One way to address the above issues is to build a single representational structure that describes all existing decision models, or, in other words, an *ontology* of decision models. Such an ontology would specify the relationships between models, allowing researchers to quantitatively measure the similarities of models, and determine whether or not the results obtained using a given model can be attributed to others. By formalizing the relationships between models, the ontology would also measure the relative flexibility of models, including both their generality (ability to mimic the predictions of other models) and their uniqueness (ability to make predictions that cannot be mimicked by others). By relating model similarities and dissimilarities to various model properties, a model ontology could also be used to test which of the mathematical assumptions in the models give rise to their idiosyncratic predictions.

Our goal in this paper is to build such an ontology of decision models. In order to do so, we perform a computational analysis that uses Monte Carlo methods to measure the relationships between models. Specifically, we calculate (potentially asymmetric) similarities and dissimilarities between pairs of models through *landscaping analysis* (Navarro, Pitt, & Myung, 2004), which measures how well one model can fit the data generated by another (this method is also sometimes called *data-uninformed parameter-bootstrapping cross-fitting*, Wagenmakers, Ratcliff, Gomez, & Iverson, 2004). We use landscaping with a wide range of randomly sampled model parameters and choice questions in order to uncover the pairwise similarities between numerous different models. Finally, various statistical and computational tools, such as multi-dimensional scaling and graph-theoretic analysis, are applied to these pairwise similarities, to interpret and analyze the representational structure captured in our ontology.

Our approach is inspired by the insights of Broomell, Budescu, & Por (2011), who use pairwise comparisons of models to understand structures of model relationships, and Pachur, Suter and Hertwig (2017), who try to integrate prospect theory and heuristic approaches to studying risky choice by analyzing model mimicry. It is also related to a number of recent papers that attempt to synthesize existing findings on choice behavior in a single representational structure or typology (Chapman et al., 2018; Eisenberg et al., 2019; Hollands et al., 2017; Norris et al., 2019). Unlike most prior work, our analysis is quantitative, and based on an established statistical technique with well-known theoretical properties. We apply this technique on a very large scale, in order to construct ontologies that include nearly every risky and intertemporal model that can be specified using a tractable parameterized mathematical function or algorithm, and subsequently use our ontologies to answer a wide range of metatheoretical questions involving model relationships and structures in decision making research.

Methods

Models

Our analysis involves 62 prominent models of risky choice and 27 prominent models of intertemporal choice, from numerous academic disciplines, published from the 1950s to the present day. We consider only mathematically tractable and parameterized models, and thus exclude general axiomatic models, qualitative (verbal) models, and simulation-based models. Details of the models are presented in Tables A1-A5 in Appendix.

Although our set of models is highly diverse, we can simplify our analysis and better interpret the model ontology by categorizing the models into a small set of discrete categories. For risky choice, we consider four core model categories: (1) Subjective expected utility theories (SEUT), which multiply transformed or untransformed payoffs against transformed or untransformed probabilities (e.g., cumulative prospect theory, Tversky & Kahneman, 1992); (2) Risk-as-value models, which explicitly incorporate a disutility caused by the riskiness (or variability) of the gamble (e.g., portfolio theory, Markowitz, 1952); (3) Counterfactual models, which compare the payoffs of gambles against alternate payoffs of the same gamble or other gambles (e.g., regret theory, Loomes & Sugden, 1982); and (4) Heuristic models, which use cognitive shortcuts to choose between gambles (e.g., the priority heuristic, Brandstätter et al., 2006).

For intertemporal choice, we consider three categories: (1) Delay discounting models, which weigh payoffs as a function of their respective delays independently (e.g., Samuelson, 1937); (2) Interval discounting models, which weigh payoffs as a function of both their delays and the interval (i.e. the difference of delays) between options (e.g., Kable & Glimcher, 2010); and (3) time-as-attribute models, which represent time delay as a separate attribute, and combine delays and payoffs using various linear and non-linear combination rules or heuristic shortcuts (e.g., Scholten & Read, 2010).

In addition to the above categories, we also analyze the underlying mathematical assumptions made by the various models. We consider four such assumptions for risky models: Whether or not the model involves (1) payoff transformations, (2) probability transformations, (3) interactions between the components (payoffs or probabilities) of a single option (“intra-option interaction), or (4) interactions between the components across options (“inter-option interaction”). The first two of these assumptions play a crucial role in the SEUT category, but also characterize many counterfactual models (which may, for example, involve a non-linear regret function applied to payoffs). The third assumption is common across all four categories of models. By allowing the outcomes of a gamble to influence the evaluation of other outcomes of the same gamble, this assumption allows a model to account for independence violations such as the Allais paradox (Allais, 1953; Kahneman & Tversky, 1979). The fourth assumption is typically only present in counterfactual models and heuristic models. By allowing the outcomes of a gamble to influence the evaluation of outcomes of other gambles, this assumption is necessary for a model to account for transitivity violations (Tversky, 1969).

We also consider three such assumptions in intertemporal choice: Whether or not the model assumes (1) non-linear transformations of delays, (2) interactions between the delays of different options, and (3) interactions between the payoffs of different options. Again, the first assumption is common in multiple model categories. The next two assumptions, which allow for the magnitude of discounting or the evaluation of payoffs of a given option to depend on other options in the choice set, can give rise to transitivity violations. The assumptions studied here are not mutually exclusive and many models apply two or more assumptions simultaneously to compute utility.

Stochastic specifications

We apply the decision models to binary choices between gambles or payoffs sequences. As most of these models are deterministic, we need to assume some type of stochastic specification in model implementation. For utility-based models, we use both Logit and Probit choice rules. The Logit choice rule defines the probability of choosing option X in a binary choice between X and Y as $p[X;Y] = \frac{1}{1+\exp\{-\varepsilon(U(X)-U(Y))\}}$, where $p[X;Y]$ is increasing in $U(X) - U(Y)$, and $1/\varepsilon$ represents the noisiness of the decision process. The larger the value of $1/\varepsilon$, the smaller the effect of $U(X) - U(Y)$ on $p[X;Y]$. Likewise, Probit defines the probability of choosing X as $p[X;Y] = \Phi\left(\varepsilon\left(U(X) - U(Y)\right)\right)$, where $0 \leq \Phi(\cdot) \leq 1$ is the cumulative standard normal distribution. In the main text we only present the results of the Logit analysis. The results of the Probit analysis can be found in supplementary materials.

The above stochastic specifications can only be applied to models that generate cardinal utilities or decision propensities. For heuristic models, which do not assign cardinal values to options, we assume a constant-error choice rule (also known as tremble noise). This stochastic specification transforms binary deterministic responses, such as a choice of X or Y, into choice probabilities $p[X;Y]$ by permitting a fixed probability $\frac{\mu}{2}$ of making an error response (with $0 \leq \mu \leq 1$). Thus, for example, if the model predicts the choice of *X*, we have $p[X;Y] = 1 - \frac{\mu}{2}$ and $p[Y;X] = \frac{\mu}{2}$.

Most existing applications of utility-based models use Logit (or Probit) stochastic specifications, and most existing applications of heuristic models use constant-error stochastic specifications. These are thus the stochastic specifications that we focus on in the main text. However, the use of different stochastic specifications for different classes of models may introduce artificial differences in model predictions, and thus distort our results. To control for this possibility we present additional analysis using only the constant-error specifications for both utility-based and heuristic models, in the supplemental materials.

Experimental designs and decision stimuli

A set of choice pairs or decision stimuli is required for the decision models to make predictions. As experimental design could be crucial in determining the (dis)similarity between models' predictions, we consider two different designs (a *main* and an *alternative* design) for generating decision stimuli for the risky and intertemporal decision domains and establishing the generalizability of the results. For risky models, our main design uses two types of binary choice questions. One type of question involves choices between two two-branch gambles, denoted as *X* = ($*x*, *p*; $0, 1-*p*) and *Y* = ($*y*, *q*; $0, 1-*q*). The other type of choice question involves choices between a sure payoff and a two-branch gamble, denoted as *X* = ($*x*, 1; $0, 0) and *Y* = ($*y*, *q*; $0, 1-*q*). There are 50 questions per choice type, totaling 100 choice questions in each choice set from the main experimental design. Our alternative design, in contrast, involves only choices between two two-branch gambles, and thus contains 100 choice questions between *X* = ($*x*, *p*; $0, 1-*p*) and *Y* = ($*y*, *q*; $0, 1-*q*).

The main design in intertemporal choice also uses two types of binary choice questions. One type of question involves choices between two delayed payoffs, that is, choices between X = (\$$x$, t) and Y = (\$$y$, s), in which $0 < x < y$ and $0 < t < s$. The other involves choices between an immediate and a delayed payoff, that is, choices between X = (\$$x$, 0) and Y = (\$$y$, s). Again, there are 50 questions of each type, totaling 100 questions in each choice set from the main experimental design. The alternative design uses only delayed payoffs, and has 100 choice questions between X = (\$$x$, t) and Y = (\$$y$, s), in which $0 < x < y$ and $0 < t < s$.

The main designs for both risky and intertemporal choice involve questions in which one choice option offers a certain or immediate payoff. We explicitly include these questions as many decision models make special predictions in the presence of certainty or immediacy. The alternative designs, in contrast, only consider a single type of question, and are thus useful for checking the robustness of our ontology in settings in which the choice set isn't specifically engineered to involve certainty or immediacy. We present the results of the main design in our main text, and present the results of the alternative design in the supplemental materials.

We generate risky and intertemporal choice questions according to the above designs by randomly sampling payoffs, probabilities, and time delays from uniform distributions. In risky choice between X = (\$$x$, p; \$0, 1-p) and Y = (\$$y$, q; \$0, 1-q), x and y are randomly and independently sampled from a uniform distribution U(0, 100); p and q are randomly, independently sampled from a uniform distribution U(0, 1). Likewise, in risky choice questions between X = (\$$x$, 1; \$0, 0) and Y = (\$$y$, q; \$0, 1-q), y is randomly sampled from uniform distribution U(0, 100) and x is randomly selected from uniform distribution U(0, y); q is randomly sampled from a uniform distribution U(0, 1).

In intertemporal choice between X = (\$$x$, t) and Y = (\$$y$, s), in which $0 < x < y$ and $0 < t < s$, y and s are randomly sampled from a uniform distribution U(0, 100); x is randomly sampled from the uniform distribution U(0, y); and t is randomly sampled from the uniform distribution U(0, s). In choices between X = (\$$x$, 0) and Y = (\$$y$, s), in which $0 < x < y$ and $0 < s$, y and s are randomly sampled from a uniform distribution U(0, 100) and x is randomly sampled from the uniform distribution U(0, y). In the alternative experimental designs all choice questions are sampled in the same manner as the first type of choice questions in the main experimental design (X = (\$$x$, 1; \$0, 0) vs. Y = (\$$y$, q; \$0, 1-q) for risky models, and X = (\$$x$, t) vs. Y = (\$$y$, s) for intertemporal models).

Note that the above stimuli involve only positive payoffs, i.e. the gain domain. However, understanding the differences between positive and negative payoffs, that is, the gain and loss domains, has been the focus of a lot of theoretical and empirical work in risky choice (e.g. Kahneman & Tversky, 1979). We focus our analysis on the gain domain as only a few risky models (mostly variants of prospect theory) explicitly differentiate between gains and losses. Most other models are explicitly formulated only for gains. Some can be made to predict loss domain phenomena, such as loss aversion, with additional assumptions not made by the initial authors (e.g. different model parameters for positive and negative payoffs), whereas others are mathematically restricted to the gain domain. That said, we present an additional set of tests using mixed gambles composed of both positive and negative payoffs in the supplemental materials. To make our risky decision models applicable to the loss domain we make some important changes to model specifications and exclude certain models from the analysis. Our results from the mixed gamble analysis are thus not directly comparable to the results for the gain domain presented in the main text.

Landscaping analysis

As discussed in the introduction, we obtain a (potentially asymmetric) measure of similarity between pairs of models by means of landscaping analysis (Navarro et al., 2004; also see Wagenmakers et al., 2004 for a related approach). Here we write the set of *N* binary choice questions as an experimental design *Q*. A generating model *G* can be written as a function f_G that takes experimental design *Q* as input and, based on a set of its parameters θ_G, produces an *N*-length vector of choice probabilities $f_G(Q|\theta_G)$ as an output. Landscaping calculates how well a second fitted model *F* is able to approximate this vector of choice probabilities. This involves searching the parameter space of *F* for some set of parameters θ_F that minimizes the dissimilarity between $f_F(Q|\theta_F)$ and $f_G(Q|\theta_G)$. We use Kullback-Leibler (KL) divergence to measure dissimilarity, and thus minimize KL divergence between $f_F(Q|\theta_F)$ and $f_G(Q|\theta_G)$, denoted as $\mathrm{D_{KL}}[f_G(Q|\theta_G) \parallel f_F(Q|\theta_F)]$. Minimizing KL divergence is equivalent to maximizing the likelihood with an infinite number of observed choice data, and using minimum KL divergence bypasses the need for simulating noisy choices numerous times to obtain accurate fit statistics. This gives our approach a degree of computational tractability not possible using standard model simulation and fitting techniques using likelihood values as a measure of fitting quality.

We implement landscaping in four steps. First, a set of $N = 100$ choice questions, *Q*, is generated in accordance with the pre-specified experimental design (outlined above). Second, for a given generating model *G*, and a given experimental design *Q*, a set of parameter values are sampled from a reasonable prior distribution (these distributions are summarized in Table A5 in Appendix). Third, *G*, with the sampled parameter values, is applied to the set of choice questions, *Q*, resulting in a 100-length vector of choice probabilities $f_G(Q|\theta_G)$. Fourth, another model, *F*, is fit to the *N*-length vector of choice probabilities by minimizing the Kullback-Leibler

(KL) divergence between the predictions of the fitted model and the predictions of the generating model. We write this measure of KL divergence as:

$$\mathrm{D_{KL}}[f_G(Q|\theta_G) \parallel f_F(Q|\theta_F)] = \sum_{q=1}^{N}\left(\sum_{o\in\{X_q,Y_q\}} f_G(o|\theta_G)\log_2\left(\frac{f_G(o|\theta_G)}{f_F(o|\theta_F)}\right)\right)$$

where $f_G(o|\theta_G)$ is the scalar predicted probability of choosing option *o* in choice question *q* (either X_q or Y_q) given model *G* and parameters θ_G. $f_F(o|\theta_F)$ is the scalar predicted probability of choosing option *o* in choice question *q* given model *F* and parameters θ_F. The summation $\sum_{o\in\{X_q,Y_q\}}(\cdot)$ measures the KL divergence of using *F* to mimic *G* for the pair of options in each choice question *q*. $\sum_{q=1}^{N}(\cdot)$ follows the chain rule of KL divergence which states that the total KL divergence over all choice questions is the sum of the KL divergences for individual choice questions.

We search for the minimum KL divergence via the Nelder-Mead simplex algorithm, implemented by MATLAB's *fminsearch* command. Here we repeat the optimization procedure in *fminsearch* 500 times with random starting points to ensure that we reach the global minimum for each fit. To ensure that all KL divergences are tractable, we constrain $f_G(o|\theta_G)$ and $f_F(o|\theta_F)$ to have a floor of 0.001 and a ceiling of 0.999. This allows us to avoid the extreme choice probabilities of 0 or 1 (for which KL divergence can be infinite). We use base-2 logarithms for calculating the KL divergences. Thus, the resulting KL divergences are in bits.

The 100 samples of *Q* and θ_G, and subsequent fits of model *F* to *G*, are used to calculate an expectation of the minimum KL divergence, which we write as:

$$d_{GF} = \mathbb{E}_{Q,\theta_G}\min_{\theta_F}\{\mathrm{D_{KL}}[f_G(Q|\theta_G) \parallel f_F(Q|\theta_F)]\}$$

d_{GF} captures how closely *F* can mimic the predictions by *G* with $d_{GF} = 0$ indicating that *F* can fit *G* perfectly. This measure is asymmetric, as one model may be able to fit the predictions generated by another, but not vice versa. Thus, we calculate d_{GF} separately for each possible combination of generating and fitted model.

As mentioned earlier, we consider three stochastic specifications (Logit, Probit and constant-error) for utility-based models and two experimental designs (a main and an alternative design) for both risky and intertemporal choice models. There are 62 risky decision models. Thus, for each combination of stochastic specification and experimental design, 3,844 (i.e. 62 × 62) pairwise model dissimilarities are estimated, resulting in a 62 × 62 asymmetric matrix. Likewise, there are 27 intertemporal decision models. Thus, for each combination of stochastic specification and experimental design, 729 (i.e. 27 × 27) model dissimilarities are estimated, resulting in a 27 × 27 asymmetric matrix. We also consider a mixed gamble design for a subset of 56 risky choice models (with the Logit stochastic specification), resulting in 56 × 56 = 3,136 pairwise model dissimilarities. As each measure of dissimilarity is approximated using 100 different samples of decision stimuli and parameters of the generating model, our entire project involves the estimation of a total of 3,057,400 minimum KL divergences (2,620,000 for risky models and 437,400 for intertemporal models). The results of the Logit/main design combination are presented in the main text. Detailed results from other combinations are presented in the supplementary materials.

Results

Reliability and generalizability

We began by testing the reliability and the generalizability of the measured model dissimilarities. We tested the former using split-half reliability. Here we divided each set of the 100 random samples for estimating d_{GF} into two halves and calculated the expectation of model dissimilarities for each half, d_{GF50}, with 50 in the subscript indicating the number of simulations in each half. We then estimated the similarity between the two halves using inner-product matrix correlation (Ramsay, ten Berge, & Styan, 1984) and implemented it using the *MatrixCorrelation* package in R (Indahl, Næs, & Liland, 2018; R Core Team, 2018). Across all the 12 computational analyses (2 choice domains × 2 experimental designs × 3 stochastic specifications), the matrix correlation coefficients between the two subsets were constantly close to 1, suggesting extremely high reliability of our measurement of model dissimilarities (see Table 1 for reliability statistics).

<Insert Table 1 about here>

We next examined the similarities of the dissimilarity matrices from different experimental designs and stochastic specifications for a test of generalizability. This was again done with inner-product matrix correlation. For both choice domains, we obtained six dissimilarity matrices, by crossing two experimental designs and three stochastic specifications. For risky decision models, these correspond to six 62×62 matrices. For intertemporal decision models, these correspond to six 27×27 matrices. We calculated the matrix correlation coefficients for each pair of matrices for each choice domain respectively. Table 1 presents the inner-product matrix correlation coefficients across different stochastic specifications and experimental designs.

Given a stochastic specification (Logit, Probit or Constant-error), the dissimilarity matrices from different experimental designs were highly consistent with each other, with all correlation coefficients above or close to 0.95 (i.e. the figures in boldface in Table 1). Turning to stochastic specifications, with the same experimental design (either Main or Alternative), Logit and Probit specifications were almost identical to each other, with correlation coefficients close to 1 for both risky and intertemporal models. This likely reflects the fact that these two stochastic specifications generate similar mappings of cardinal utility to choice probability. Even with different experimental designs, the correlation coefficients between Logit and Probit always exceed 0.93 for both risky and intertemporal models.

The correlation coefficients between Logit/Probit and the constant-error stochastic specifications are, however, slightly lower. As shown in Table 1, these range between 0.79 and 0.88 for risky decision models, and between 0.75 and 0.84 for intertemporal choice models, depending on the experimental design. These results indicate that stochastic specification can influence a model's quantitative predictions (see Blavatskyy & Pogrebna 2010; Loomes & Sugden, 1995; Regenwetter et al., 2018; Scholten, Read & Sanborn, 2014 for extended discussion). Nonetheless the correlations are all fairly high, indicating that our dissimilarity matrices, and subsequent ontologies, are fairly stable.

Note that we also analyzed an experimental design with both gains and losses in the risky choice domain. These tests resulted in somewhat different model dissimilarity matrices. However, these results are not directly comparable to those from the gain domain presented above, as our extension to the loss domain required fundamental changes to model specifications and the exclusion of a subset of risky choice models. We elaborate on these differences in the supplemental materials.

Model spaces

The set of model dissimilarities obtained through landscaping quantify model relationships for each pair of models. To better interpret these relationships, we used the pairwise similarities to drive representations of the models as points in a multidimensional space. Such spatial representations provide an intuitive description of similarities across numerous models. Additionally, central points in such spaces identify prototypical models and peripheral points in such spaces identify atypical and unusual models, allowing for an intuitive understanding of the representational structure captured in the model dissimilarity matrices.

In order to obtain spatial representations, we first symmetrized our measures of model dissimilarity: $\overline{d_{GF}} = \overline{d_{FG}} = \frac{d_{GF}+d_{FG}}{2}$. We projected these symmetrized model dissimilarity measures onto latent dimensions via non-metric multidimensional scaling (NMDS) (Kruskal, 1964; Venables & Ripley, 2002). The non-metric approach relaxes the assumption of a cardinal distance measure of classical multidimensional scaling and relies solely on the rank order of the symmetrized KL divergence, $\overline{d_{GF}}$. Thus, the NMDS solutions would hold constant even if any other distance measure that is monotonically increasing in $\overline{d_{GF}}$ is used. We obtained NMDS representations by minimizing the stress of the low-dimensional configurations (Kruskal, 1964). To ensure that the global minimum stress was reached, the optimization procedure was repeated 100,000 times with random starting configurations.

Two-dimensional representations of the space of risky and intertemporal models, obtained through the above methods applied to the Logit stochastic specification and the main experimental design, are shown in Figures 1a and 2a respectively. Analogous figures for alternative stochastic specifications and designs are provided in Figures S1 and S3 of supplemental materials, and the relationships between model spaces derived using different stochastic specifications and experimental designs is summarized in Tables S1 and S2 of supplemental materials. The figures also show the centroid of the spaces using black crosses. Here we represent different categories of our risky models (SEUT, risk-as-value, counterfactual, heuristic) and intertemporal models (delay discounting, interval discounting, time-as-attribute) using different colors.

<Insert Figure 1 about here>

The model space in Figure 1a illustrates the latent structure of risky decision models, and can be used to identify theoretical distinctions that result in diverging model predictions and subsequently large model distances. For example, SEUT models tend to cluster with each other in the central region of the model space. This is likely due to the fact that subjective transformations to payoffs and probabilities are among the earliest and most influential assumptions in modeling risky choice. We observe some clustering within the risk-as-value category and the counterfactual category, whose models occupy most of the left region of the model space in Figure 1a. The proximity of these two sets of models may be due to the fact that some disappointment-based counterfactual models closely resemble risk-as-value models, as they implicitly place a penalty on high variance gambles. We can also see that heuristic models are mostly located at the periphery of the space. These models make extreme predictions that diverge sharply from those made through utility maximization.

There are also interesting types of variation within each category of models. For example, although most SEUT models are located at the center of the space, a few are located at the periphery. These include expected value maximization (#1), subjective expected money (#4), certainty equivalence theory (#5) and dual theory (#8 and #9). None of these peripheral models involve transformations to payoffs; they all assume linear value functions.

Likewise, although all risk-as-value models share a similar model structure (which aggregates the moments of the gamble's distribution), some have very idiosyncratic predictions, such as the mean-variance-skewness model (#27). In contrast, others, such as the alpha-target model (#28), the below-target model (#29), the below-mean semivariance model (#31), and the coefficient-of-variation model (#37), are very similar to expected value maximization. These differences are likely the result of the ways in which gamble moments enter the utility function, and how this relates to the design choices in our analysis.

There is a fair amount of spread in the locations of counterfactual models. These locations appear to rely more on whether a payoff transformation is applied in the model, rather than the mechanism the model represents (e.g. regret vs disappointment). This is why, for example, regret theory with expected value evaluation and disappointment theory with expected value evaluation (#39 and #42) are located close to each other and to the expected value maximization model (#1), whereas regret theory with expected utility evaluation and disappointment theory with expected utility evaluation (#40 and #43) are located close to the expected utility maximization model (#2).

Finally, there is a significant amount of variability in the location of the heuristic models. A few heuristic models with operations that involve utility-based calculation, such as the low expected payoff elimination heuristic (#50), the relative expected loss immunization heuristic (#58) and the similarity heuristic with expected utility evaluation (#60), are located closer to the cluster of utility-based models. Generally, however, there are many important differences in the types of predictions made by different heuristics, which is why the heuristics are spread out over a relatively large region.

We can also examine the structure of the model space by categorizing models based on their specific mathematical assumptions rather than their broad theoretical interpretations. This is done in Figure S2 of supplemental materials, which categorizes risky models based on whether they involve non-linear payoff transformations, probability transformations, both, or neither. This figure suggests that non-linear transformations, especially payoff transformation, play a key role in determining the similarities and differences between risky choice models. Models that apply both payoff and probability transformations are located close to each other and to the centre of the model space. Models with payoff transformations (but without probability transformations) are nearby. By contrast, models without these transformations are located at the periphery of the space. We return to this point in the next section of this paper.

<Insert Figure 2 about here>

In intertemporal choice, we observe a large distinction between discounting models and time-as-attribute models (Figure 2a), as they involve fundamental differences in the representation and valuation of time. We also observe a distinction between delay discounting and interval discounting models, though all discounting models are relatively close to each other, and clustered near the center of the model space. In fact, contrary to prior work that focuses on distinctions between exponential and hyperbolic discounting, we find that these two model classes are quite similar to each other. For example, the one-parameter hyperbolic discounting model (#2) is closer to the (one-parameter) exponential discounting model (#1), than it is to other hyperbolic discounting models. This is not to say that exponential and hyperbolic models are identical; they can of course be distinguished by manually crafting an appropriate stimulus set (Read, 2001; Green, Myerson & Macaux, 2005; Scholten & Read, 2006) or algorithmically sampling the most discriminative stimuli (e.g., Cavagnaro et al., 2016). However, with randomly generated stimuli, as in the current analysis, these two classes of models make quite similar predictions.

Time-as-attribute models, in contrast, are spread over a large area in the periphery of the model space, reflecting their idiosyncratic predictions and properties. Indeed, some of these models are considered heuristics and heuristics in risky choice are also often spread out over the periphery of the model space. Note that among time-as-attribute models, ITCH (#26) produces qualitative predictions similar to tradeoff models (see discussion in Ericson et al. 2015). However, our landscaping analysis suggests that its quantitative predictions can be quite distinct from the latter's (#22, #24, and #25). These three tradeoff models are all located close to each other, and are also relatively close to discounting models, indicating that some time-as-attribute models can approximate discounting (see discussion in Scholten et al., 2014).

Property cohesion

The high degree of clustering observed for SEUT models in Figure 1a suggests that using a multiplicative combination of (often transformed) payoffs and probabilities plays an important role in the models' predictions. In this sense, belonging or not belonging to the SEUT category is a critical property of a decision model, and determines the model's position in our ontology. Pairs of models that both belong to the SEUT category can mimic each other and are positioned close to each other. If one model belongs to the SEUT category and the other doesn't, the models are unlikely to be able to mimic each other or be near each other in our space. Figure 2a shows a similar degree of proximity for models that fall within the delay discounting category, suggesting that delay discounting is a similarly critical property.

These claims can be made more rigorously by measuring the cohesion of each category, or, more generally, each property associated with a model. For each model property p, we define its cohesion coefficient c_p as the difference between the average dissimilarity across all models and the average dissimilarity between models sharing the property:

$$c_p = \text{AVE}\{d_{GF} | G, F \in M_{All} \text{ and } G \neq F\} - \text{AVE}\{d_{GF} | G, F \in M_p \text{ and } G \neq F\},$$

where M_{All} represents the full set of decision models within a choice domain and M_p represents the subset of models that have property p. This cohesion coefficient is larger for model properties that, if shared by two models, are likely to result in proximate positions in the model space (with the two models making similar predictions and being able to closely mimic each other's predictions). Intuitively these are properties that have the strongest effects on model predictions, and play the largest role in differentiating model predictions from each other.

We calculated the above cohesion coefficients for the various model properties discussed in the Methods section, including model category (e.g. SEUT, risk-as-value etc.) and underlying mathematical assumptions of the model (e.g. payoff transformations, probability transformations etc.). We also estimated confidence bounds of the cohesion coefficients via permutation. In this calculation, we randomly selected a set of models of the same size as the set size of M_p, and calculated a hypothetical cohesion coefficient for the set. This was repeated 100,000 times to estimate the two-tail 95% confidence bounds on the distribution of the permutation-based cohesion coefficients. Cohesion coefficients that lie above or below these bounds can be seen to be significantly positive or negative, that is, unlikely to arise by chance.

Cohesion coefficients and confidence bounds are displayed in Figure 3a for risky models and Figure 3b for intertemporal models. A positive cohesion coefficient indicates that sharing the corresponding property results in convergent model predictions while a negative cohesion coefficient indicates divergent model predictions. As in Figures 1a and 2a, we can see that SEUT and delay discounting models are highly coherent categories, with significantly positive cohesion coefficients. Risk-as-value, counterfactual, and interval discounting models are also somewhat coherent. Heuristic risky models and time-as-attribute models, in contrast, are highly incoherent categories, containing dissimilar models that yield diverging predictions.

<Insert Figure 3 about here>

We likewise see that the mathematical assumptions that has the largest positive cohesion coefficients is payoff transformation for risky models and delay transformation for intertemporal models. For risky models, probability transformation also seems to produce high cohesion coefficients. In contrast, assumptions about the interactions between various choice components have non-significant (and sometimes negative) cohesion coefficients. Thus, it seems that non-linear transformations of payoffs and probabilities for risky models and delays for intertemporal models play a crucial role in determining a model's predictions. Models that share these assumptions are likely to be able to closely mimic each other. In this sense, these transformations are essential properties of the models – they determine the positions of the models in the model ontology.

Additionally, different disciplines and different historical time periods may have different degrees of coherence, capturing historical patterns of paradigm development, and an analysis of cohesion coefficients for models belonging to different disciplines and time periods can provide a valuable meta-scientific perspective on decision modeling. For risky models, we observe consistently positive and significant cohesion coefficients in management and economics, suggesting that risky models published within these disciplines are often highly similar to each other (perhaps a consequence of paradigm consensus in these fields). Psychology models, in contrast, are significantly dissimilar to each other, with cohesion coefficients below the negative confidence bound. This could be due to the fact that psychology admits numerous, often divergent perspectives, and that many non-utility models (such as heuristics and cognitive computational models) are published in psychology journals. We do not observe obvious historical trends in model cohesion, though the 1980s appear to be significantly incoherent and the 1990s appear to be significantly coherent. We speculate that these trends could be due to publication of many heuristics in the 1980s, and the publication of many SEUT models (including numerous variants of prospect theory) in the 1990s.

For intertemporal models, we also observe positive cohesion coefficients for models published in management and neuroscience journals, although these properties do not reach the significance threshold. This is likely due to the fact that there are only a few models from these two disciplines, resulting in wide confidence bounds. Conversely, models published in economics and psychology journals have coherence coefficients around zero. This could be because these disciplines have seen the publication of a variety of different types of intertemporal models. Finally, we observe significantly positive cohesion coefficients for models published in 1990s and earlier, indicating that early intertemporal decision models are similar to each other. In contrast, we observe negative or low cohesion coefficients for models published in 2000s and 2010s. The decline in model cohesion over time is likely driven by the surging interest in the attribute-based views of intertemporal choice in recent decades.

The property cohesion coefficients shown in Figure 3 can be influenced by the choice of models used in the analysis. For example, there may be changes to these coefficients if we removed multiple variants of prospect theory from our model space. However, we believe that the use of all models (including multiple variants of prospect theory) is more informative, as these variants are typically proposed by different researchers, and published in different papers at different points in time. The multiplicity of prospect theory models is, in this sense, part of the status quo in risky decision modeling, and thus should be reflected in the results of our analysis.

Finally note that Figure 3 shows property cohesion coefficients for the analysis involving Logit stochastic specifications applied to the main experimental design. Analogous results with alternate stochastic specifications and experimental designs are provided in Figures S4 and S5 of supplemental materials.

Directed graphs

Another representation for our model ontology involves directed graphs (Chartrand, 1977). Unlike the spaces analyzed above, graphs have the benefit of accommodating asymmetries in model dissimilarities, which are measures of model hierarchy (if one model can mimic another, but not vice versa, the second model can be seen as a restricted version of the first). To obtain such graphs, we discretized model dissimilarities, so that model *F* has a connection to model *G* (i.e. fits the data generated by *G*), if d_{GF} is smaller than some threshold value. We visualize these graphs in Figure 1b for risky models, and Figure 2b for intertemporal models, using a threshold of $d_{GF} = 0.011$ (which from an information-theoretic perspective corresponds to *F*'s best-fit predictions having at least 95% overlap with *G*'s generated choice probabilities (for an illustration, see Broomell & Bhatia, 2014, as well as our discussion in supplementary materials). Again these graphs involve only the analysis using the Logit stochastic specification applied to the main experimental design. Analogous graphs with alternate stochastic specifications and experimental designs are provided in Figures S6 and S7 of supplemental materials. Table S2 in supplemental materials summarizes the relationship between the graphs in Figures 1 and 2 and those in Figures S6 and S7.

Figures 1b and 2b allow us to examine the relationships between pairs of models to test if the behavior of one model can be described by another. The node size in the graphs is proportional to the model's total connectedness to other models (i.e., the sum of outgoing and incoming connections). These figures reveal a number of interesting relationships between models. Expected value and expected utility theories are the most connected models in risky choice, and exponential discounting is the most connected model in intertemporal choice. These models are mimicked by a number of different types of models, including models in other categories (e.g. exponential discounting is mimicked by some interval discounting models).

Prominent behavioral models like cumulative prospect theory are also capable of mimicking the behavior of other types of models, such as the minimax heuristic, which assumes that people choose gambles that offer the highest minimum payoffs.

We also observe a number of cliques in our graphs. Cliques are sets of models that are all mutually connected to each other, and thus all give similar predictions to each other. The largest such clique involves expected value maximization and four risk-as-value models, which all appear to mimic expected value maximization in our tests. Other cliques involve sets of prospect theory variants and sets of heuristic models in risky choice, and sets of hyperbolic models in intertemporal choice (see supplementary material Tables S3-S6 for details of model cliques).

Model generality and uniqueness

Graphs also allow us to study the overall generality and uniqueness of models. Specifically, the number of outgoing connections, or outdegree centrality, of a target model corresponds to the number of other models that can be mimicked by the target model, and is thus a measure of the generality of the target model. Likewise, the number of incoming connections, or indegree centrality, of a target model corresponds to the number of other models that are capable of mimicking data generated by the target model, and is thus a measure of the (inverse) uniqueness or idiosyncrasy of the target model. The degree centralities of the models in our analysis are provided in Tables S7 and S8 of supplemental materials.

Generality (in the form of outdegree centrality) measures a model’s ability to predict data generated by other models and uniqueness (in the form of the inverse of indegree centrality) measures its ability to generate data that other models cannot predict. For this reason, generality and uniqueness are two manifestations of relative model flexibility. As expected, these two measures depend on the number of parameters in the model, so that models with more parameters have higher outdegree centralities and lower indegree centralities. In risky choice, we observe rank correlations of 0.74 ($p < .001$) and -0.22 ($p = .08$) between the number of parameters and outdegree and indegree centrality respectively.

The number of parameters in a model is not the only determinant of its place in the model hierarchy. For example, models with the highest outdegree centrality in risky choice are three SEUT models, which include two variants of cumulative prospect theory (#12 and #17). These models have only four total parameters each (there are a total of nine risky decision models with more than four parameters). The high outdegree centrality of these models, despite their relatively small number of parameters, likely reflects the central role of this framework in guiding theoretical risky choice research: Subjective expected utility models are among the earliest behavioral models of risky choice, and many subsequent models are variants or special cases of subjective expected utility.

Correspondingly, we find that the models with the highest indegree centralities are expected value maximization (#1) and expected utility maximization (#2). Most of these models have two or more parameters, and again are not the least parameterized models (there are 16 risky models with only one free parameter). The high indegree centrality of expected value and expected utility reflects the fact that they have served as benchmark models in risky choice research, with many more complex models subsuming expected value and expected utility as special cases.

In intertemporal models, outdegree centrality depends on the number of parameters with a rank correlation of 0.60 ($p < .001$). The models with the highest outdegree centrality are an interval discounting model and three hyperbolic discounting models. The former has a very large number of parameters, and allows for delays and payoffs to be combined in many different ways.

The latter, like their subjective expected utility counterparts, were some of the earliest behavioral models of intertemporal choice. Indegree centrality depends on the number of parameters to a lesser extent, with a rank correlation of -0.12 (p = .55). The exponential discounting model (#1), for example, has the highest indegree centrality but is not among the models with the fewest number of parameters. Once again, this reflects the fact that exponential discounting has served as a benchmark model for intertemporal choice research.

Note that the measures of model flexibility analyzed here are relative measures that hold between different models. They are thus somewhat distinct from the (absolute) measure of model flexibility in the statistical model comparison framework, which defines model flexibility in terms of a model's ability to capture regions of the data space (see e.g., Pitt, Myung, & Zhang, 2002). Of course, as in the standard statistical model comparison framework, high flexibility within our framework is not a desirable feature of a model. Although it does indicate that a given model can mimic others, this is likely due to the ability of the model to predict a broad range of data. Thus models with high relative flexibility (i.e. high generality and low uniqueness) are also likely models with high absolute flexibility as defined in the statistical model comparison framework.

Discussion

Our paper has showcased a metatheoretical analysis, which is capable of quantifying the relationships between different decision models, and can be used to derive a model ontology in the form of low-dimensional spaces and directed graphs of decision models. Our ontology sheds light on the theoretical assumptions of decision models that have the strongest effects on model predictions, and which play the largest role in distinguishing models from each other. Our ontology also identifies prototypical models of choice behavior. These are models which closely resemble other models, and which subsequently lie at the centers of our spaces and graphs. Finally, our ontology allows us to characterize the hierarchical structure of models, which offers precise measurements of model generality and uniqueness.

Perhaps the most important contribution of our paper is in synthesizing mathematical and computational research on choice behavior across academic disciplines over the past 70 years. Despite its fundamental role in science and society, we currently do not have a unified computational theory of human choice behavior. In fact, the vast interdisciplinary scope and long history of decision research have resulted in over 80 different models of simple risky and intertemporal choice alone. By building an ontology of decision models, we provide a single framework within which different decision models can be represented. Such a representational framework does not only help researchers better understand the theoretical properties of and relationships between models, but also allows for the generalization of theoretical and empirical insights across research programs and academic disciplines. Thus, for example, brain regions that have been found to encode preferences corresponding to a particular model can also be assumed to encode preferences corresponding to the model's neighbors, which likely include numerous decision models not studied by neuroscientists. Likewise, socioeconomic or demographic variables that have been shown to influence the parameters of a given economic model likely also influence the parameters of proximate models, which may have been proposed by psychologists. The converse is true for affective, cognitive, and clinical variables studied using psychological models.

By analyzing the relationships between different decision models and the core features of the space of decision models, our approach complements more established techniques in decision research such as axiomatic analysis (von Neumann & Morgenstern, 1947; Fishburn, 1970; Luce

& Marley, 2005). Axiomatic analysis identifies critical qualitative conditions that give rise to general functional representations of decision models, and in turn, differentiate different decision models from each other. In contrast to this, our approach measures the relationships between different decision models based on how well they can mimic each other. Unlike axiomatic analysis, our approach can be applied to nearly any decision model, including models that do not have easily discernable axiomatic properties. It is useful to note that the two approaches do not always yield the same results. For example, Figure 3 shows that models that have inter-option interaction in risky choice (and thus violate the transitivity axiom) do not necessarily occupy neighboring positions in our model ontology. Understanding these divergences is a promising topic for future work.

Our approach is also closely related to the information geometric approach to functional form analysis in statistics (Amari, 1985). The information geometric approach sees a model as a geometric object, with each point in the object representing a distribution of model predictions. For a binary decision model *G*, each point in the geometric object represents the vector of choice probabilities over the set of decision stimuli given parameters θ_G, i.e., $f_G(Q|\theta_G)$. This approach has been applied to evaluate model complexity by estimating the volume of the geometric object that represents the model and can therefore be used for model selection (Myung et al., 2000; Pitt et al., 2002; Grünwald, 2007). Instead of estimating the volume of the geometric objects that represent models' predictions, our landscaping analysis focuses on the interaction between different models' predictions i.e. the relationship between *G*'s prediction $f_G(Q|\theta_G)$ and *F*'s predictions $f_F(Q|\theta_F)$.

Of course, our specific results depend on the set of experimental stimuli we use for the landscaping analysis. We have considered two different designs for generating stimuli, and for each design, have generated choice pairs using random draws from probability distributions over payoffs, probabilities and time delays. This is a common approach to generating stimuli in decision modeling as it ensures a diverse array of stimuli combinations (Rieskamp, 2008; Erev, Ert, Plonsky, Cohen, & Cohen, 2017), and thus leads to high levels of parameter identifiability (Broomell & Bhatia, 2014). Interestingly we find a high degree of consistency between the model ontologies generated using our two different experimental designs. This may, in part, be due to the use of random sampling in our stimuli generation process, which ensures that our stimuli vary across each sample in our Monte Carlo tests, thereby giving our tests a degree of generality not possible using a single fixed set of stimuli.

Randomly generated stimuli may not offer a good approximation to the types of choice questions decision makers encounter in the world (Pleskac & Hertwig, 2014), and, alternate experimental designs may be preferable. In fact, by quantifying model relationships, our approach allows for a formal analysis of the effect of design choice on model behavior (see Navarro et al., 2004 and Wagenmakers et al., 2004 for additional discussions of this issue; also see Pitt & Myung, 2009; Cavagnaro, Pitt, Gonzalez, & Myung, 2013; Cavagnaro, Aranovich, McClure, Pitt, & Myung, 2016). Thus, it is possible to use variants of the landscaping approach to algorithmically uncover the types of decision problems for which a given model makes unique predictions.

One example of this application is provided in our supplemental materials, where we analyze model ontologies generated with mixed gambles (see e.g. Figure S8). Here we find that models that explicitly distinguish between gains and losses are closer to each other, and further away from related models that do not explicitly distinguish between gains and losses. We are cautious about interpreting differences between these mixed-gamble ontologies and the (gain gamble) ontologies in the main text, as our mixed-gamble implementations exclude a number of models, and make important modifications to the rest. Our results may also change if we allow all models to have separate parameters for positive and negative payoffs (as is the case for the probability weighting functions in prospect theory models). In any case, the results of these preliminary tests illustrate a new type of application for the computational approach presented in this paper, and future work could extend these tests to more rigorously compare model spaces generated using gain gambles and mixed gambles. This work could also examine the effects of other common design choices, such as the exclusion of easy choices involving dominated gambles or gambles with large differences in expected value, or the oversampling of gambles with very small or very large probabilities (see e.g. Erev et al., 2002; Rieskamp, 2008). A similar type of analysis could also of course be applied to the intertemporal choice domain.

Finally, our analysis relates to recent research on theory integration in risky choice. For example, Pachur et al. (2017; see also Pachur et al., 2018) have shown that the parameters governing payoff and probability transformations in the cumulative prospect theory (CPT) can be used to mimic the predictions of decision heuristics, such as the minimax. This close link between CPT and minimax is also reflected in Figure 1b, which indicates that our approach is able to replicate Pachur et al.'s core results. Figure 1b also displays connections between minmax and many other risky models. In fact, there are a total of 291 connections between 53 unique models in this figure. This suggests that theory integration is possible on a much larger scale than that attempted by Pachur et al.

To aid this type of theory integration, we have released our set of computed pairwise model dissimilarities, two-dimensional model spaces, and directed model graphs, in the supplemental materials. We envision researchers using the model relationships derived as part of our ontology to extend theoretical and empirical claims made using individual models, to the diverse array of models that are currently studied in the behavioral sciences. We also expect future work to build off the ideas outlined in this paper, so as to advance the representational frameworks for describing theories of choice behavior. Such an endeavor could utilize actual empirical data to constrain model parameters, try to combine our model ontologies for risky and intertemporal choice, relate our ontology to decision theoretic axioms satisfied by different models, or study the effects of experimental design on the resulting ontology. Ultimately, an ontology of decision models offers a powerful theoretical framework for interpreting the numerous psychological, economic, and neurobiological correlates of choice, and is necessary for a cumulative, trans-disciplinary science of human choice behavior.

References

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque, critique des postulats et axiomes de l'ecole americaine. *Econometrica, 21*, 503-546.

Amari, S. I. (1985). *Differential-geometrical methods in statistics* (Vol. 28). Springer Science & Business Media.

Anchugina, N. (2017). A simple framework for the axiomatization of exponential and quasi-hyperbolic discounting. *Theory and Decision, 82*(2), 185-210.

Bell, D. E. (1982). Regret in decision making under uncertainty. *Operations Research*, *30*(5), 961-981.

Bell, D. E. (1985). Disappointment in decision making under uncertainty. *Operations Research*, *33*(1), 1-27.

Benhabib, J., Bisin, A., & Schotter, A. (2010). Present-bias, quasi-hyperbolic discounting, and fixed costs. *Games and Economic Behavior*, *69*(2), 205-223.

Bernoulli, D. (1738). Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, *1738*(5), 175-192. (Translated into English by L. Sommer in *Econometrica*, 1954, 22, 23-36.)

Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive consumer choice processes. *Journal of Consumer Research, 25*(3), 187-217.

Bhatia, S. (2014). Sequential sampling and paradoxes of risky choice. *Psychonomic Bulletin & Review*, *21*(5), 1095-1111.

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. (2008). New paradoxes of risky decision making. *Psychological Review, 115*(2), 463–501.

Blavatskyy, P. R. & Pogrebna, G. (2010). Models of stochastic choice and decision theories: Why both are important for analyzing decisions. *Journal of Applied Econometrics, 25*(6), 963-986.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly Journal of Economics*, *127*(3), 1243-1285.

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

Broomell, S. B., & Bhatia, S. (2014). Parameter recovery for decision modeling using choice data. *Decision, 1*(4), 252-274.

Broomell, S. B., Budescu, D. V., & Por H. H. (2011). Pair-wise comparisons of multiple models. *Judgment & Decision Making, 6*(8), 821-831.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review, 100*(3), 432-459.

Camerer, C. F., Loewenstein, G., & Rabin, M. (Eds.). (2004). *Advances in Behavioral Economics*. Princeton University Press.

Cavagnaro, D. R., Aranovich, G. J., McClure, S. M., Pitt, M. A., & Myung, J. I. (2016). On the functional form of temporal discounting: An optimized adaptive test. *Journal of Risk and Uncertainty, 52*(3), 233-254.

Cavagnaro, D. R., Pitt, M. A., Gonzalez, R., & Myung, J. I. (2013). Discriminating among probability weighting functions using adaptive design optimization. *Journal of Risk and Uncertainty, 47*(3), 255-289.

Chapman, J., Dean, M., Ortoleva, P., Snowberg, E., & Camerer, C. (2018). Econographics (No. w24931). *National Bureau of Economic Research*.

Chartrand, G. (1977). *Introductory Graph Theory*. Courier Corporation.

Cheng, J., & González-Vallejo, C. (2016). Attribute-wise vs. alternative-wise mechanism in intertemporal choice: Testing the proportional difference, trade-off, and hyperbolic models. *Decision, 3*(3), 190-215.

Coombs, C. H., & Pruitt, D. G. (1960). Components of risk in decision making: Probability and variance preferences. *Journal of Experimental Psychology, 60*(5), 265-277.

Dai, J., & Busemeyer, J. R. (2014). A probabilistic, dynamic, and attribute-wise model of intertemporal choice. *Journal of Experimental Psychology: General, 143*(4), 1489-1514.

Delquié, P., & Cillo, A. (2006). Disappointment without prior expectation: a unifying perspective on decision under risk. *Journal of Risk and Uncertainty*, *33*(3), 197-215.

Diecidue, E., & Van De Ven, J. (2008). Aspiration level, probability of success and failure, and expected utility. *International Economic Review*, *49*(2), 683-700.

Dyer, J. S., & Jia, J. (1997). Relative risk-value models. *European Journal of Operational Research*, *103*(1), 170-185.

Ebert, J. E. & Prelec, D. (2007). The fragility of time: time-insensitivity and valuation of the near and far future. *Management Science*, *53* (9), 1423–1438.

Edwards, W. (1955). The prediction of decisions among bets. *Journal of Experimental Psychology, 50*(3), 201–214

Eisenberg, I. W., Bissett, P. G., Enkavi, A. Z., Li, J., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Uncovering the structure of self-regulation through data-driven ontology discovery. *Nature communications, 10*(1), 2319.

Erev, I., Roth, A., Slonim, R. L., & Barron, G. (2002). Combining a theoretical prediction with experimental evidence to yield a new prediction: An experimental design with a random sample of tasks. *Unpublished manuscript*.

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*(4), 369-409.

Ericson, K. M., White, J. M., Laibson, D., & Cohen, J. D. (2015). Money earlier or later? Simple heuristics explain intertemporal choices better than delay discounting does. *Psychological Science, 26*(6), 826-833.

Frederick, S., Loewenstein, G., & O'donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature, 40*(2), 351-401.

Fishburn, P. C. (1970). *Utility Theory for Decision Making*. New York: Wiley.

Fishburn, P. C. (1977). Mean-risk analysis with risk associated with below-target returns. *The American Economic Review*, *67*(2), 116-126.

Glimcher, P. W., & Fehr, E. (Eds.). (2013). *Neuroeconomics: Decision Making and the Brain*. Academic Press.

Glimcher, P. W., & Rustichini, A. (2004). Neuroeconomics: the consilience of brain and decision. *Science, 306*(5695), 447-452.

Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology*, *38*(1), 129-166.

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics* (Vol. 1). New York: Wiley.

Green, L., & Myerson, J. (2004). A Discounting Framework for Choice with Delayed and Probabilistic Rewards. *Psychological Bulletin, 130*(5), 769-792.

Green, L., Myerson, J., & Macaux, E. W. (2005). Temporal discounting when the choice is between two delayed rewards. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 1121-1133.

Handa, J. (1977). Risk, probabilities, and a new theory of cardinal utility. *Journal of Political Economy*, *85*(1), 97-122.

Hogarth, R. M., & Einhorn, H. J. (1990). Venture theory: A model of decision weights. *Management Science*, *36*(7), 780-803.

Hollands, G. J., Bignardi, G., Johnston, M., Kelly, M. P., Ogilvie, D., Petticrew, M., ... & Marteau, T. M. (2017). The TIPPME intervention typology for changing environments to change behaviour. *Nature Human Behaviour, 1*(8), 0140.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review, 92*(5), 1644-1655.

Indahl, U. G., Næs, T., & Liland, K. H. (2018). A similarity index for comparing coupled matrices. *Journal of Chemometrics, 32*(10), e3049.

Kable, J. W., & Glimcher, P. W. (2010). An “as soon as possible” effect in human intertemporal decision making: behavioral evidence and neural mechanisms. *Journal of Neurophysiology, 103*(5), 2513-2531.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica, 47*(2), 263-292.

Karmarkar, U. S. (1978). Subjectively weighted utility: A descriptive extension of the expected utility model. *Organizational Behavior and Human Performance*, *21*(1), 61-72.

Killeen, P. R. (2009). An additive-utility model of delay discounting. *Psychological Review, 116*(3), 602-619.

Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika, 29*(1), 1-27.

Kruskal, J. B. (1964b). Nonmetric multidimensional scaling: a numerical method. *Psychometrika, 29*(2), 115-129.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79-86.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics, 112*(2), 443-478.

Lattimore, P. K., Baker, J. R., & Witte, A. D. (1992). The influence of probability on risky choice: A parametric examination. *Journal of Economic Behavior & Organization*, *17*(3), 377-400.

Leland, J. W. (1994). Generalized similarity judgments: An alternative explanation for choice anomalies. *Journal of Risk and Uncertainty*, *9*(2), 151-172.

Leland, J. W. (2002). Similarity judgments and anomalies in intertemporal choice. *Economic Inquiry*, *40*(4), 574-581.

Lieder, F., Griffiths, T. L., & Hsu, M. (2018). Overrepresentation of extreme events in decision making reflects rational use of cognitive resources. *Psychological Review, 125*(1), 1-32.

Loewenstein, G., O'Donoghue, T., & Bhatia, S. (2015). Modeling the interplay between affect and deliberation. *Decision, 2*(2), 55-81.

Loewenstein, G., & Prelec, D. (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *The Quarterly Journal of Economics, 107*(2), 573-597.

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

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal, 92(*368), 805-824.

Loomes, G., & Sugden, R. (1986). Disappointment and dynamic consistency in choice under uncertainty. *The Review of Economic Studies*, *53*(2), 271-282.

Loomes, G., & Sugden, R. (1995). Incorporating a stochastic element into decision theories. *European Economic Review, 39*(3-4), 641-648.

Luce, R. D., & Marley, A. A. (2005). Ranked additive utility representations of gambles: Old and new axiomatizations. *Journal of Risk and Uncertainty, 30*(1), 21-62.

MacKillop, J., Amlung, M. T., Few, L. R., Ray, L. A., Sweet, L. H., & Munafò, M. R. (2011). Delayed reward discounting and addictive behavior: a meta-analysis. *Psychopharmacology, 216*(3), 305-321.

Marchiori, D., Di Guida, S., & Erev, I. (2015). Noisy retrieval models of over- and undersensitivity to rare events. *Decision, 2*(2), 82-106.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance, 7*(1), 77-91.

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In M. L. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.), *The effect of delay and of intervening events on reinforcement value: quantitative analyses of behavior* (pp. 55–73). Hillsdale, NJ: Lawrence Erlbaum.

McClure, S. M., Ericson, K. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2007). Time discounting for primary rewards. *Journal of Neuroscience, 27*(21), 5796-5804.

McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science, 306*(5695), 503-507.

Mellers, B., Schwartz, A., & Ritov, I. (1999). Emotion-based choice. *Journal of Experimental Psychology: General*, *128*(3), 332.

Mukherjee, K. (2010). A dual system model of preferences under risk. *Psychological Review, 117*(1), 243–255.

Myung, I. J., Balasubramanian, V., & Pitt, M. A. (2000). Counting probability distributions: Differential geometry and model selection. *Proceedings of the National Academy of Sciences, 97*(21), 11170-11175.

Myung, J. I., & Pitt, M. A. (2009). Optimal experimental design for model discrimination. *Psychological Review, 116*(3), 499–518.

Navarro, D. J., Pitt, M. A., & Myung, I. J. (2004). Assessing the distinguishability of models and the informativeness of data. *Cognitive Psychology, 49*(1), 47-84.

Pachur, T., Schulte-Mecklenbeck, M., Murphy, R. O., & Hertwig, R. (2018). Prospect theory reflects selective allocation of attention. *Journal of Experimental Psychology: General,* 147, 147-169.

Pachur, T., Suter, R. S., & Hertwig, R. (2017). How the twain can meet: Prospect theory and models of heuristics in risky choice. *Cognitive Psychology*, 93, 44-73.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 534-552.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review, 109*(3), 472-291.

Pleskac, T. J., & Hertwig, R. (2014). Ecologically rational choice and the structure of the environment. *Journal of Experimental Psychology: General, 143*(5), 2000-2019.

Prelec, D. (1998). The Probability Weighting Function. *Econometrica*, *66*(3), 497-528.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rachlin, H. (2006). Notes on discounting. *Journal of the Experimental Analysis of Behavior*, *85*(3), 425-435.

Ramsay, J. O., ten Berge, J., & Styan, G. P. H. (1984). Matrix correlation. *Psychometrika*, *49*(3), 403-423.

Read, D. (2001). Is time-discounting hyperbolic or subadditive? *Journal of Risk and Uncertainty*, *23*(1), 5–32.

Read, D., Frederick, S., & Scholten, M. (2013). DRIFT: An analysis of outcome framing in intertemporal choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(2), 573-588.

Regenwetter, M., Cavagnaro, D. R., Popova, A., Guo, Y., Zwilling, C., Lim, S. H., & Stevens, J. R. (2018). Heterogeneity and parsimony in intertemporal choice. *Decision, 5*(2), 63-94.

Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(6), 1446-1465.

Roelofsma, P. H. (1996). Modelling intertemporal choices: An anomaly approach. *Acta Psychologica*, *93*(1-3), 5-22.

Rubinstein, A. (1988). Similarity and decision-making under risk (Is there a utility theory resolution to the Allais paradox?). *Journal of Economic Theory*, *46*(1), 145-153.

Samuelson, P. A. (1937). A note on measurement of utility. *The Review of Economic Studies, 4*(2), 155-161.

Savage, Leonard J. 1954. *The Foundations of Statistics*. New York, Wiley.

Scholten, M., & Read, D. (2006). Discounting by intervals: A generalized model of intertemporal choice. *Management Science, 52*(9), 1424-1436.

Scholten, M., & Read, D. (2010). The psychology of intertemporal tradeoffs. *Psychological Review, 117*(3), 925-944.

Scholten, M., Read, D., & Sanborn, A. (2014). Weighing outcomes by time or against time? Evaluation rules in intertemporal choice. *Cognitive Science*, *38*(3), 399–438.

Sibson, R. (1978). Studies in the robustness of multidimensional scaling: Procrustes statistics. *Journal of the Royal Statistical Society: Series B (Methodological)*, *40*(2), 234-238.

Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature, 38*(2), 332-382.

Thorngate, W. (1980). Efficient decision heuristics. *Behavioral Science*, *25*(3), 219-225.

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

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*(4), 297-323.

Tversky, A., & Kahneman, D. (Eds.). (2000). *Choices, Values, and Fames*. Cambridge University Press.

Venables, W. N., & Ripley, B. D. (2013). *Modern Applied Statistics with S-PLUS*. Springer Science & Business Media.

Viscusi, W. K. (1989). Prospective reference theory: Toward an explanation of the paradoxes. *Journal of Risk and Uncertainty*, *2*(3), 235-263.

Von Neumann, J. & Morgenstern, O. (1947). *Theory of Games and Economic Behavior*. Princeton University Press.

Wagenmakers, E. J., Ratcliff, R., Gomez, P., & Iverson, G. J. (2004). Assessing model mimicry using the parametric bootstrap. *Journal of Mathematical Psychology, 48*(1), 28-50.

Weber, E. U., & Johnson, E. J. (2009). Mindful judgment and decision making. *Annual Review of Psychology, 60*, 53-85.

Weber, E. U., Shafir, S., & Blais, A.-R. (2004). Predicting Risk Sensitivity in Humans and Lower Animals: Risk as Variance or Coefficient of Variation. *Psychological Review, 111*(2), 430-445.

Wu, G., Zhang, J., & Abdellaoui, M. (2005). Testing prospect theories using probability tradeoff consistency. *Journal of Risk and Uncertainty*, *30*(2), 107-131.

Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica: Journal of the Econometric Society*, 95-115.

Table 1

Reliability and generalizability of the measure of model dissimilarities via inner-product matrix correlation coefficients.

			Risky choice						Intertemporal choice					
			Main			Alternative			Main			Alternative		
			L	P	C	L	P	C	L	P	C	L	P	C
Reliability			0.99	0.99	0.93	0.99	1.00	0.98	0.99	0.99	0.97	0.98	0.98	0.97
Generalizability	Main	L	1						1					
		P	1.00	1					1.00	1				
		C	0.87	0.88	1				0.81	0.82	1			
	Alternative	L	**0.96**	0.95	0.79	1			**0.94**	0.93	0.75	1		
		P	0.96	**0.96**	0.81	1.00	1		0.95	**0.95**	0.77	1.00	1	
		C	0.83	0.84	**0.95**	0.82	0.83	1	0.83	0.84	**0.97**	0.79	0.81	1

Note. L = Logit; P = Probit; C=Constant-error.

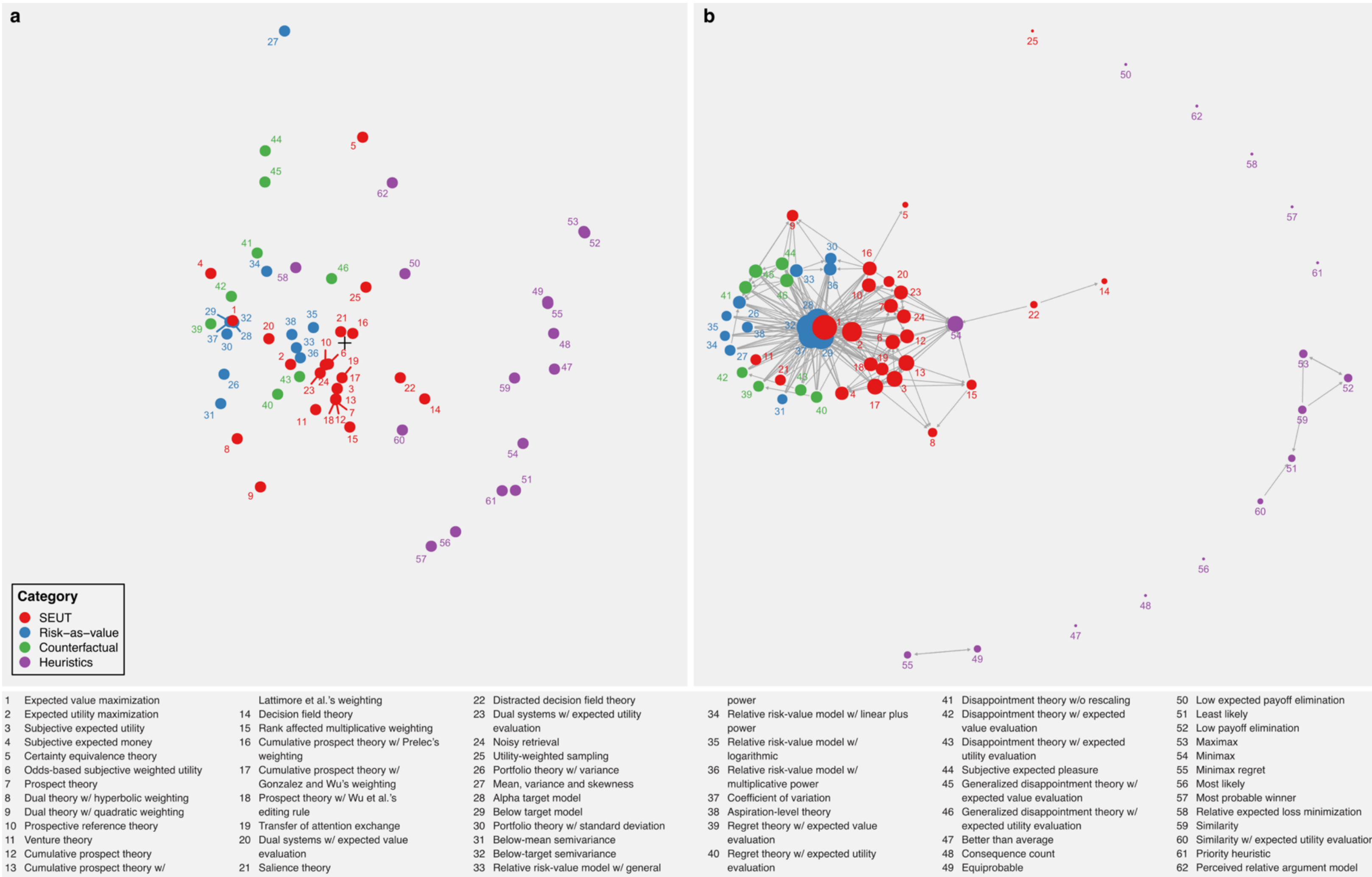

Figure 1. Ontology of risky decision models. (a) Two-dimensional model space from non-metric multidimensional scaling, with a black cross representing the centroid of the space. (b) Directed graph representation, with node size corresponding to the total connectedness of a model (i.e. the sum of indegree and outdegree centralities). The full list of risky models is provided at the bottom of Figure 1.

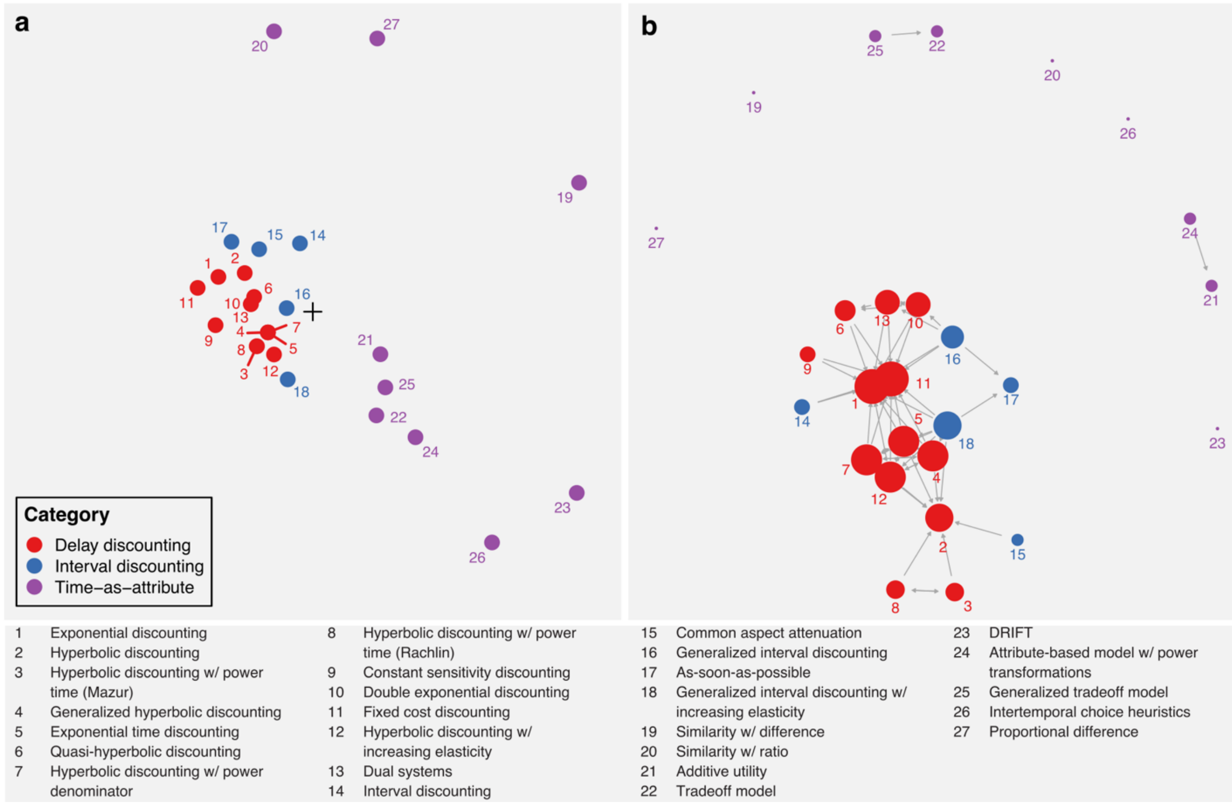

*Figure 2***.** Ontology of intertemporal decision models. (a) Two-dimensional model space from non-metric multidimensional scaling, with a black cross representing the centroid of the space. (b) Directed graph representation, with node size corresponding to the total connectedness of a model (i.e. the sum of indegree and outdegree centralities). The full list of intertemporal models is provided at the bottom of Figure 2.

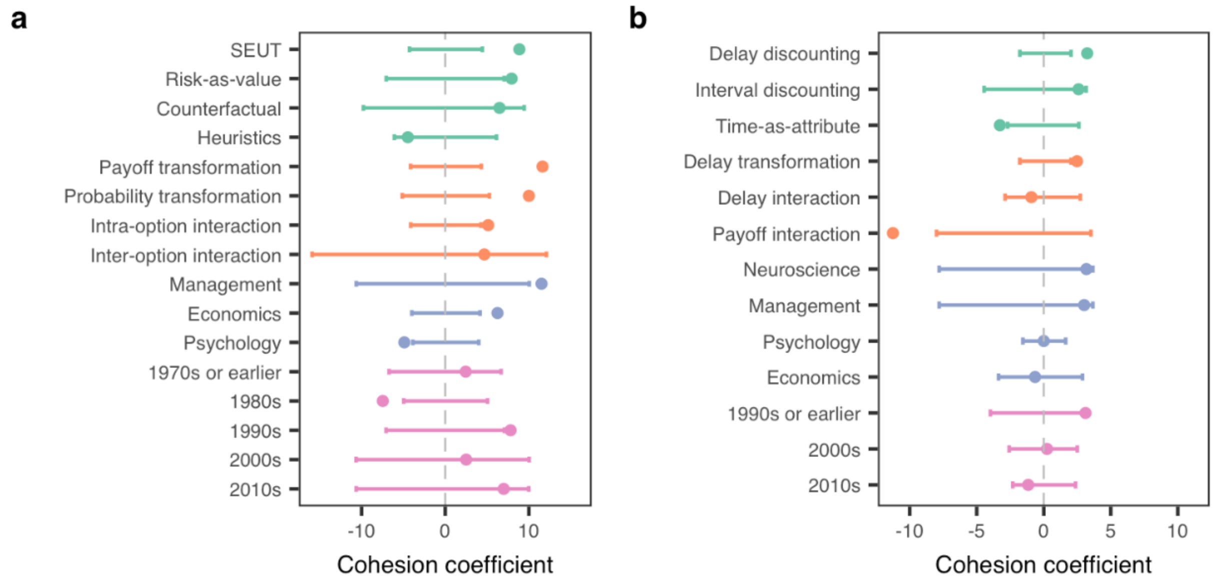

Figure 3. Cohesion coefficients for model properties**.** Error bars represent the permutation-based 95% confidence bounds at chance level. (**a**) Risky decision models and (**b**) intertemporal decision models.

Appendix: Decision Models and Mathematical Functions

Table A1

Summary of risky decision models.

ID	Model	Category	Authors	Year	Source (Journal or Book)
1	Expected value maximization	SEUT			
2	Expected utility maximization	SEUT	Bernoulli	1738	Papers of the Imperial Academy of Sciences in Petersburg
3	Subjective expected utility	SEUT	Savage	1954	The Foundations of Statistics
4	Subjective expected money	SEUT	Edwards	1955	Journal of Experimental Psychology
5	Certainty equivalence theory	SEUT	Handa	1977	Journal of Political Economy
6	Odds-based subjective weighted utility	SEUT	Karmarkar	1978	Organizational Behavior and Human Performance
7	Prospect theory	SEUT	Kahneman and Tversky	1979	Econometrica
8	Dual theory w/ hyperbolic weighting	SEUT	Yaari	1987	Econometrica

9	Dual theory w/ quadratic weighting	SEUT	Yaari	1987	Econometrica
10	Prospective reference theory	SEUT	Viscusi	1989	Journal of Risk and Uncertainty
11	Venture theory	SEUT	Hogarth and Einhorn	1990	Management Science
12	Cumulative prospect theory	SEUT	Tversky and Kahneman	1992	Journal of Risk and Uncertainty
13	Cumulative prospect theory w/ Lattimore et al.'s weighting	SEUT	Lattimore et al.	1992	Journal of Economic Behavior and Organization
14	Decision field theory	SEUT	Busemeyer and Townsend	1993	Psychological Review
15	Rank affected multiplicative weighting	SEUT	Birnbaum	1997	Choice, Decision, and Measurement: Essays in Honor of R. Duncan Luce
16	Cumulative prospect theory w/ Prelec's weighting	SEUT	Prelec	1998	Econometrica
17	Cumulative prospect theory w/ Gonzalez and Wu's weighting	SEUT	Gonzalez and Wu	1999	Cognitive Psychology

18	Prospect theory w/ Wu et al.'s editing rule	SEUT	Wu et al.	2005	Journal of Risk and Uncertainty
19	Transfer of attention exchange	SEUT	Birnbaum	2008	Psychological Review
20	Dual systems w/ expected value evaluation	SEUT	Mukherjee	2010	Psychological Review
21	Salience theory	SEUT	Bordalo et al.	2012	Quarterly Journal of Economics
22	Distracted decision field theory	SEUT	Bhatia	2014	Psychonomic Bulletin & Review
23	Dual systems w/ expected utility evaluation	SEUT	Loewenstein et al.	2015	Decision
24	Noisy retrieval	SEUT	Marchiori et al.	2015	Decision
25	Utility-weighted sampling	SEUT	Lieder et al.	2018	Psychological Review
26	Portfolio theory w/ variance	Risk-as-value	Markowitz	1952	Journal of Finance
27	Mean, variance and skewness	Risk-as-value	Coombs and Pruitt	1960	Journal of Experimental Psychology
28	Alpha target model	Risk-as-value	Fishburn	1977	American Economic Review
29	Below target model	Risk-as-value	Fishburn	1977	American Economic Review

30	Portfolio theory w/ standard deviation	Risk-as-value	Fishburn	1977	American Economic Review
31	Below-mean semivariance	Risk-as-value	Fishburn	1977	American Economic Review
32	Below-target semivariance	Risk-as-value	Fishburn	1977	American Economic Review
33	Relative risk-value model w/ general power	Risk-as-value	Dyer and Jia	1997	European Journal of Operational Research
34	Relative risk-value model w/ linear plus power	Risk-as-value	Dyer and Jia	1997	European Journal of Operational Research
35	Relative risk-value model w/ logarithmic	Risk-as-value	Dyer and Jia	1997	European Journal of Operational Research
36	Relative risk-value model w/ multiplicative power	Risk-as-value	Dyer and Jia	1997	European Journal of Operational Research
37	Coefficient of variation	Risk-as-value	Weber	2004	Psychological Review
38	Aspiration-level theory	Risk-as-value	Diecidue and van de Ven	2008	International Economic Review

39	Regret theory w/ expected value evaluation	Counterfactual	Bell	1982	Operations Research
40	Regret theory w/ expected utility evaluation	Counterfactual	Loomes and Sugden	1982	Economic Journal
41	Disappointment theory w/o rescaling	Counterfactual	Bell	1985	Operations Research
42	Disappointment theory w/ expected value evaluation	Counterfactual	Loomes and Sugden	1986	Review of Economic Studies
43	Disappointment theory w/ expected utility evaluation	Counterfactual	Loomes and Sugden	1986	Review of Economic Studies
44	Subjective expected pleasure	Counterfactual	Mellers et al.	1999	Journal of Experimental Psychology: General
45	Generalized disappointment theory w/ expected value evaluation	Counterfactual	Delquié and Cillo	2006	Journal of Risk and Uncertainty
46	Generalized disappointment theory w/ expected utility evaluation	Counterfactual	Delquié and Cillo	2006	Journal of Risk and Uncertainty
47	Better than average	Heuristics	Thorgate	1980	Behavioral Science

48	Consequence count	Heuristics	Thorgate	1980	Behavioral Science
49	Equiprobable	Heuristics	Thorgate	1980	Behavioral Science
50	Low expected payoff elimination	Heuristics	Thorgate	1980	Behavioral Science
51	Least likely	Heuristics	Thorgate	1980	Behavioral Science
52	Low-payoff elimination	Heuristics	Thorgate	1980	Behavioral Science
53	Maximax	Heuristics	Thorgate	1980	Behavioral Science
54	Minimax	Heuristics	Thorgate	1980	Behavioral Science
55	Minimax Regret	Heuristics	Thorgate	1980	Behavioral Science
56	Most likely	Heuristics	Thorgate	1980	Behavioral Science
57	Most probable winner	Heuristics	Thorgate	1980	Behavioral Science
58	Relative expected loss minimization	Heuristics	Thorgate	1980	Behavioral Science
59	Similarity	Heuristics	Rubinstein	1988	Journal of Economic Theory
60	Similarity w/ expected utility evaluation	Heuristics	Leland	1994	Journal of Risk and Uncertainty
61	Priority heuristic	Heuristics	Brandstatter et al.	2006	Psychological Review

62	Perceived relative argument model	Heuristics	Loomes	2010	Psychological Review

Table A2

Functional forms of risky decision models. The notations are designed for choices between $X = (\$x_1, p_1;\ \$x_2, p_2)$ *and* $Y = (\$y_1, q_1;\ \$y_2, q_2)$, *where* $x_1 > x_2$,. $y_1 > y_2$, $p_1 + p_2 = 1$ *and* $q_1 + q_2 = 1$. $U(X)$ *denotes the utility or choice propensity of X and* $U(Y)$ *denotes the utility or choice propensity of Y. If not given,* $U(Y)$ *can be obtained by replacing* x_i *and* p_i *with* y_i *and* q_i *respectively in* $U(X)$. *When* $U(X)$ *involves interactions with* y_i *or* q_i, *the corresponding* $U(Y)$ *replaces* y_i *and* q_i *in* $U(X)$ *with* x_i *and* p_i *respectively. For heuristic models,* $A(X)$ *denotes the argument for option X and* $A(Y)$ *denotes the argument for option Y. If not given,* $A(Y)$, *can be obtained by replacing* x_i *and* p_i *with* y_i *and* q_i *respectively. When* $A(X)$ *involves interactions with* y_i *or* q_i, *the corresponding* $A(Y)$ *should replace* y_i *and* q_i *with* x_i *and* p_i *respectively. Free parameters are denoted by Greek letters, with corresponding domains and prior distributions shown in Table A5.*

ID	**Model**	**Function**	**Stochastic specification**[1]
1	Expected value maximization	$U(X) = \sum_{i=1}^{2} p_i x_i$	Logit, Probit or Constant-error
2	Expected utility maximization	$U(X) = \sum_{i=1}^{2} p_i u(x_i)$	Logit, Probit or Constant-error

[1] NA is the abbreviation of "not applicable", meaning that the model itself involves a stochastic specification.

3	Subjective expected utility	$U(X) = \sum_{i=1}^{2} w_{LBW}(p_i)u(x_i)$	Logit, Probit or Constant-error
4	Subjective expected money	$U(X) = \sum_{i=1}^{2} w_K(p_i)x_i$	Logit, Probit or Constant-error
5	Certainty equivalence theory	$U(X) = \sum_{i=1}^{2} \sqrt{p_i}x_i$	Logit, Probit or Constant-error
6	Odds-based subjective weighted utility	$U(X) = \sum_{i=1}^{2} \frac{w_K(p_i)u(x_i)}{\sum_{j=1}^{2} w_K(p_j)}$	Logit, Probit or Constant-error
7	Prospect theory	$U(X) = \begin{cases} w_{TK}^{+}(p_1)\big(u_{PT}(x_1) - u_{PT}(x_2)\big) + u_{PT}(x_2), \text{if } x_1 > 0 \\ u_{PT}(x_1) + w_{TK}^{-}(p_2)\big(u_{PT}(x_2) - u_{PT}(x_1)\big), \text{if } x_1 \leq 0 \end{cases}$	Logit, Probit or Constant-error
8	Dual theory w/ hyperbolic weighting	$U(X) = x_2 + \frac{p_1}{2 - p_1}(x_1 - x_2)$	Logit, Probit or Constant-error
9	Dual theory w/ quadratic weighting	$U(X) = x_2 + {p_1}^2(x_1 - x_2)$	Logit, Probit or Constant-error
10	Prospective reference theory	$U(X) = \sum_{i=1}^{2} \left((1-\mu)p_i + \frac{\mu}{2}\right)u(x_i)$	Logit, Probit or Constant-error

11	Venture theory	$U(X) = \sum_{i=1}^{2} w_V(p_i)u(x_i)$	Logit, Probit or Constant-error
12	Cumulative prospect theory	$U(X) = \begin{cases} w_{TK}^{+}(p_1)u_{PT}(x_1) + \left(1 - w_{TK}^{+}(p_1)\right)u_{PT}(x_2), & \text{if } 0 \leq x_2 \leq x_1 \\ w_{TK}^{+}(p_1)u_{PT}(x_1) + w_{TK}^{-}(p_2)u_{PT}(x_2), & \text{if } x_2 < 0 < x_1 \\ \left(1 - w_{TK}^{-}(p_2)\right)u_{PT}(x_1) + w_{TK}^{-}(p_2)u_{PT}(x_2), & \text{if } x_2 \leq x_1 \leq 0 \end{cases}$	Logit, Probit or Constant-error
13	Cumulative prospect theory w/ Lattimore et al.'s weighting	$U(X) = \begin{cases} w_{LBW}^{+}(p_1)u_{PT}(x_1) + \left(1 - w_{LBW}^{+}(p_1)\right)u_{PT}(x_2), & \text{if } 0 \leq x_2 \leq x_1 \\ w_{LBW}^{+}(p_1)u_{PT}(x_1) + w_{LBW}^{-}(p_2)u_{PT}(x_2), & \text{if } x_2 < 0 < x_1 \\ \left(1 - w_{LBW}^{-}(p_2)\right)u_{PT}(x_1) + w_{LBW}^{-}(p_2)u_{PT}(x_2), & \text{if } x_2 \leq x_1 \leq 0 \end{cases}$	Logit, Probit or Constant-error
14	Decision field theory	$p[X;Y] = \frac{1}{1+\exp\{-\epsilon \cdot d\}}$, where $d = \frac{2\left(\sum_{i=1}^{2} p_i u(x_i) - \sum_{j=1}^{2} q_j u(y_j)\right)}{p_1 p_2 \left(u(x_1) - u(x_2)\right)^2 + q_1 q_2 \left(u(y_1) - u(y_2)\right)^2}$	NA
15	Rank affected multiplicative weighting	$U(X) = \sum_{i=1}^{2} \frac{i\, {p_i}^{\tau}}{\sum_{j=1}^{2} j\, {p_j}^{\tau}} u(x_i)$	Logit, Probit or Constant-error
16	Cumulative prospect theory w/ Prelec's weighting	$U(X) = \begin{cases} w_{P}^{+}(p_1)u_{PT}(x_1) + \left(1 - w_{P}^{+}(p_1)\right)u_{PT}(x_2), & \text{if } 0 \leq x_2 \leq x_1 \\ w_{P}^{+}(p_1)u_{PT}(x_1) + w_{P}^{-}(p_2)u_{PT}(x_2), & \text{if } x_2 < 0 < x_1 \\ \left(1 - w_{P}^{-}(p_2)\right)u_{PT}(x_1) + w_{P}^{-}(p_2)u_{PT}(x_2), & \text{if } x_2 \leq x_1 \leq 0 \end{cases}$	Logit, Probit or Constant-error

17	Cumulative prospect theory w/ Gonzalez and Wu's weighting	$U(X) = \begin{cases} w_{GW}^{+}(p_1)u_{PT}(x_1) + \left(1 - w_{GW}^{+}(p_1)\right)u_{PT}(x_2), \text{if } 0 \leq x_2 \leq x_1 \\ w_{GW}^{+}(p_1)u_{PT}(x_1) + w_{GW}^{-}(p_2)u_{PT}(x_2), \quad \text{if } x_2 < 0 < x_1 \\ \left(1 - w_{GW}^{-}(p_2)\right)u_{PT}(x_1) + w_{GW}^{-}(p_2)u_{PT}(x_2), \text{if } x_2 \leq x_1 \leq 0 \end{cases}$	Logit, Probit or Constant-error
18	Prospect theory w/ Wu et al.'s editing rule	$U(X) = \begin{cases} w_{TK}^{+}(p_1)u_{PT}(x_1 - x_2) + u_{PT}(x_2), \text{if } 0 < x_1 \\ u_{PT}(x_1) + w_{TK}^{-}(p_2)u_{PT}(x_2 - x_1), \text{if } x_1 \leq 0 \end{cases}$	Logit, Probit or Constant-error
19	Transfer of attention exchange[2]	$U(X) = \frac{\left({p_1}^{\tau} - \frac{\kappa}{3}{p_2}^{\tau}\right)u(x_1) + \left({p_2}^{\tau} + \frac{\kappa}{3}{p_1}^{\tau}\right)u(x_2)}{{p_1}^{\tau} + {p_2}^{\tau}}$	Logit, Probit or Constant-error
20	Dual systems w/ expected value evaluation	$U(X) = (1 - \mu)\sum_{i=1}^{2} p_i x_i + \mu \frac{\sum_{i=1}^{2} u(x_i)}{2}$	Logit, Probit or Constant-error
21	Salience theory	$U(X) = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{\beta^{r_{ij}} p_i q_j}{\sum_{k=1}^{2}\sum_{l=1}^{2} \beta^{r_{kl}} p_k q_l} u(x_i)$, where r_{ij} corresponds to the rank order of the four pairwise salience values $s_{ij} = \frac{\lvert x_i - y_j \rvert}{x_i + y_j + \alpha + \rho \cdot \mathbb{1}(x_i + y_j \geq 0)}$.	Logit, Probit or Constant-error
22	Distracted decision field theory	$p[X;Y] = \frac{1}{1+\exp\{-\epsilon \cdot d\}}$, where	NA

[2] This is a "special" TAX model assuming that all weight transfers are the same fixed proportion of the branch giving up weight (Birnbaum, 2008; pp. 470, Eq.8a).

		$d = 2\left(\sum_{i=1}^{2}\left(\left((1-\mu)p_i + \frac{\mu}{2}\right)u(x_i)\right) - \sum_{j=1}^{2}\left(\left((1-\mu)q_j + \frac{\mu}{2}\right)u(y_j)\right)\right) \div \left(\left(\prod_{i=1}^{2}\left((1-\mu)p_i + \frac{\mu}{2}\right)\right)\left(u(x_1) - u(x_2)\right)^2 + \prod_{j=1}^{2}\left((1-\mu)q_j + \frac{\mu}{2}\right)\left(u(y_1) - u(y_2)\right)^2\right)$	
23	Dual systems w/ expected utility evaluation	$U(X) = \sum_{i=1}^{2} p_i u(x_i) + \kappa \sum_{i=1}^{2}(\mu + (1-\mu)p_i)u(x_i)$	Logit, Probit or Constant-error
24	Noisy retrieval	$U(X) = \sum_{i=1}^{2}\left((1-\mu)p_i + \frac{\mu}{2}\right)u(x_i)$	Logit, Probit or Constant-error
25	Utility-weighted sampling[3]	$p[X;Y] = (1-\mu)\left(\Pr\left(k > \frac{\iota}{2}; r, \iota\right) + \frac{1}{2}\Pr\left(k = \frac{\iota}{2}; r, \iota\right)\right) + \frac{\mu}{2}$, where $r = \frac{\sum_{i=1}^{2}\sum_{j=1}^{2} \mathrm{I}\left(u_R(x_i) \geq u_R(y_j)\right)\lvert u_R(x_i) - u_R(y_j)\rvert p_i q_j}{\sum_{i=1}^{2}\sum_{j=1}^{2}\lvert u_R(x_i) - u_R(y_j)\rvert p_i q_j}$ is the probability of sampling X from $\{X, Y\}$ and $\Pr(\cdot)$ is the binomial probability mass function of sampling X for k times out of the total ι times, the latter of which is a free parameter for the model.	NA

[3] This is a simplification of the original (simulation based) utility-weighted sampling theory.

26	Portfolio theory w/ variance	$U(X) = \sum_{i=1}^{2} p_i x_i - \kappa\, p_1 p_2\, (x_1 - x_2)^2$	Logit, Probit or Constant-error
27	Mean, variance and skewness	$U(X) = \sum_{i=1}^{2} p_i x_i + \nu\, p_1 p_2\, (x_1 - x_2)^2 + \phi \frac{p_2 - p_1}{\sqrt{p_1 p_2}}$	Logit, Probit or Constant-error
28	Alpha target model[4]	$U(X) = \sum_{i=1}^{2} p_i x_i - \kappa \sum_{i=1}^{2} \mathrm{I}(x_i < 100\delta)\, p_i (100\delta - x_i)^{\alpha}$	Logit, Probit or Constant-error
29	Below target model	$U(X) = \sum_{i=1}^{2} p_i x_i - \kappa \sum_{i=1}^{2} \mathrm{I}(x_i < 100\delta)\, p_i (100\delta - x_i)$	Logit, Probit or Constant-error
30	Portfolio theory w/ standard deviation	$U(X) = \sum_{i=1}^{2} p_i x_i - \kappa \sqrt{p_1 p_2}\, (x_1 - x_2)$	Logit, Probit or Constant-error
31	Below-mean semivariance	$U(X) = \sum_{i=1}^{2} p_i x_i - \kappa\, p_2 \left(\sum_{i=1}^{2} p_i x_i - x_2 \right)^2$	Logit, Probit or Constant-error
32	Below-target semivariance	$U(X) = \sum_{i=1}^{2} p_i x_i - \kappa \sum_{i=1}^{2} \mathrm{I}(x_i < 100\delta) p_i (100\delta - x_i)^2$	Logit, Probit or Constant-error

[4] $100\,\delta$ (with $0 \leq \delta \leq 1$) represents the target value in this model and other target-related models.

33	Relative risk-value model w/ general power	$U(X)=\left(\sum_{i=1}^{2}p_ix_i\right)^{\tau}+\beta\left(\sum_{i=1}^{2}p_ix_i\right)^{1-\lambda}\left(\sum_{j=1}^{2}p_j\left(\frac{x_j}{\sum_{i=1}^{2}p_ix_i}\right)^{\theta}-1\right)$	Logit, Probit or Constant-error
34	Relative risk-value model w/ linear plus power	$U(X)=\sum_{i=1}^{2}p_ix_i-\beta\left(\sum_{i=1}^{2}p_ix_i\right)^{1-\lambda}\left(\sum_{j=1}^{2}p_j\left(\frac{x_j}{\sum_{i=1}^{2}p_ix_i}\right)^{1-\kappa}-\nu\right)$	Logit, Probit or Constant-error
35	Relative risk-value model w/ logarithmic	$U(X)=\frac{1}{\alpha}\log\left(1+\alpha\sum_{i=1}^{2}p_ix_i\right)+\frac{\kappa}{\alpha}\sum_{j=1}^{2}p_j\log\left(1+\alpha\frac{x_j}{\sum_{i=1}^{2}p_ix_i}\right)$	Logit, Probit or Constant-error
36	Relative risk-value model w/ multiplicative power	$U(X)=\left(\sum_{i=1}^{2}p_ix_i\right)^{\tau}\sum_{j=1}^{2}p_j\left(\frac{x_j}{\sum_{i=1}^{2}p_ix_i}\right)^{\theta}$	Logit, Probit or Constant-error
37	Coefficient of variation	$U(X)=\sum_{i=1}^{2}p_ix_i-\kappa\,\frac{\sqrt{p_1p_2}\,(x_1-x_2)}{100\,\sum_{i=1}^{2}p_ix_i}$	Logit, Probit or Constant-error
38	Aspiration-level theory	$U(X)=\sum_{i=1}^{2}p_iu(x_i)+\alpha\sum_{i=1}^{2}p_i\,\mathrm{I}(x_i\geq k)-\beta\sum_{i=1}^{2}p_i\,\mathrm{I}(x_i<k)$, where $k=\min\{x_2,y_2\}+\delta(\max\{x_1,y_1\}-\min\{x_2,y_2\})$ is the aspiration level.	Logit, Probit or Constant-error
39	Regret theory w/ expected value evaluation	$U(X)=\sum_{i=1}^{2}p_ix_i+\sum_{i=1}^{2}\sum_{j=1}^{2}p_i\,q_j\mathrm{R}(x_i-y_j)$	Logit, Probit or Constant-error

40	Regret theory w/ expected utility evaluation	$U(X)=\sum_{i=1}^{2}p_i u(x_i)+\sum_{i=1}^{2}\sum_{j=1}^{2}p_i\,q_j \mathrm{R}\left(u(x_i)-u(y_j)\right)$	Logit or Probit
41	Disappointment theory w/o rescaling	$U(X)=\sum_{i=1}^{2}p_i x_i+\nu\,p_1 p_2(x_1-x_2)$	Logit, Probit or Constant-error
42	Disappointment theory w/ expected value evaluation	$U(X)=\sum_{i=1}^{2}p_i x_i+\kappa\sum_{j=1}^{2}p_i\ \mathrm{sign}\left(x_j-\sum_{i=1}^{2}p_i x_i\right)\left\lvert x_j-\sum_{i=1}^{2}p_i x_i\right\rvert^{\alpha}$	Logit, Probit or Constant-error
43	Disappointment theory w/ expected utility evaluation	$U(X)=\sum_{i=1}^{2}p_i u(x_i) +\kappa\sum_{j=1}^{2}p_j\ \mathrm{sign}\left(u(x_j)-\sum_{i=1}^{2}p_i u(x_i)\right)\left\lvert u(x_j) -\sum_{i=1}^{2}p_i u(x_i)\right\rvert^{\alpha}$	Logit, Probit or Constant-error
44	Subjective expected pleasure	$U(X)=\sum_{i=1}^{2}p_i x_i+p_1 p_2\left(\rho(x_1-x_2)^{\chi}-\upsilon(x_1-x_2)^{\psi}\right) +\sum_{i=1}^{2}\sum_{j=1}^{2}p_i\,q_j \mathrm{R}\left(u(x_i)-u(y_j)\right)$	Logit, Probit or Constant-error

45	Generalized disappointment theory w/ expected value evaluation	$U(X) = \sum_{i=1}^{2} p_i x_i + p_1 p_2 (\kappa(x_1 - x_2)^{\alpha} - \lambda(x_1 - x_2)^{\beta})$	Logit, Probit or Constant-error
46	Generalized disappointment theory w/ expected utility evaluation	$U(X) = \sum_{i=1}^{2} p_i u(x_i) + p_1 p_2 \left(\kappa\big(u(x_1) - u(x_2)\big)^{\alpha} - \lambda\big(u(x_1) - u(x_2)\big)^{\beta}\right)$	Logit, Probit or Constant-error
47	Better than average	$A(X) = \sum_{i=1}^{2} \mathrm{I}\left(x_i > \frac{1}{4}(x_1 + x_2 + y_1 + y_2)\right)$	Constant-error
48	Consequence count	$A(X) = \sum_{i=1}^{2} \mathrm{sign}(x_i - y_i)$	Constant-error
49	Equiprobable	$A(X) = \frac{1}{2}\sum_{i=1}^{2} x_i$	Constant-error
50	Low expected payoff elimination	$A(X) = 2\ \mathrm{sign}(p_1 x_1 - q_1 y_1) + \mathrm{sign}(p_2 x_2 - q_2 y_2)$	Constant-error
51	Least likely	$A(X) = p_1$	Constant-error
52	Low-payoff elimination	$A(X) = 2\ \mathrm{sign}(x_1 - y_1) + \mathrm{sign}(x_2 - y_2)$	Constant-error
53	Maximax	$A(X) = \mathrm{sign}(x_1 - y_1)$	Constant-error

54	Minimax	$A(X) = \text{sign}(x_2 - y_2)$	Constant-error
55	Minimax Regret	$A(X) = \min(x_1 - y_1, x_2 - y_2)$	Constant-error
56	Most likely	$A(X) = \sum_{i=1}^{2} \left(\frac{1}{2} + \frac{1}{2}\,\text{sign}\left(p_i - \frac{1}{2}\right)\right) x_i$	Constant-error
57	Most probable winner	$A(X) = \sum_{i=1}^{2} \sum_{j=1}^{2} p_i q_j \mathrm{I}(x_i > y_j)$	Constant-error
58	Relative expected loss minimization	$p[X;Y] = \frac{1}{2} + \kappa \frac{2(A(X) - A(Y))}{A(X) + A(Y)}$, where $A(X) = \sum_{i=1}^{2} \sum_{j=1}^{2} p_i q_j \min(x_i - y_j, 0)$	NA
59	Similarity	$A(X) = \text{sign}(x_1 - y_1)\,\mathrm{I}\left(\frac{\min(x_1, y_1)}{\max(x_1, y_1)} \leq \delta\right) + \text{sign}(p_1 - q_1)\,\mathrm{I}\left(\frac{\min(p_1, q_1)}{\max(p_1, q_1)} \leq \tau\right)$	Constant-error

60	Similarity w/ expected utility evaluation	$A(X) = \Bigg(\mathrm{I}\left(\sum_{i=1}^{2} p_i u(x_i) - \sum_{j=1}^{2} q_j u(y_j) > \alpha \right) - \mathrm{I}\left(\sum_{j=1}^{2} q_j u(y_j) - \sum_{i=1}^{2} p_i u(x_i) > \alpha \right) \Bigg) \times 4 + \mathrm{sign}\big(\mathrm{sign}(x_2 - y_2) + \mathrm{I}(x_2 = y_2)\mathrm{sign}(q_2 - p_2) + \mathrm{I}(x_1 \geq y_1)\mathrm{I}(p_1 \geq q_1) - \mathrm{I}(x_1 \leq y_1)\mathrm{I}(p_1 \leq q_1)\big) \times 2 + \mathrm{sign}\Big(\mathrm{I}(x_2 - y_2 \geq \beta) - \mathrm{I}(y_2 - x_2 \geq \beta) + \mathrm{I}(\lvert x_2 - y_2 \rvert < \beta)\,\mathrm{sign}(q_2 - p_2)\,\mathrm{I}\left(\lvert q_2 - p_2 \rvert \geq \frac{\delta}{2}\right) + \mathrm{I}(x_1 - y_1 > -\beta)\,\mathrm{I}\left(p_1 - q_1 > -\frac{\delta}{2}\right) - \mathrm{I}(x_1 - y_1 < \beta)\,\mathrm{I}\left(p_1 - q_1 < \frac{\delta}{2}\right)\Big)$	Constant-error
61	Priority heuristic	$A(X) = \mathrm{sign}(x_2 - y_2)\,\mathrm{I}\left(\lvert x_2 - y_2 \rvert > \frac{\max(x_1, y_1)}{10}\right) \times 4 + \mathrm{sign}(p_1 - q_1)\,\mathrm{I}(\lvert p_1 - q_1 \rvert > 0.1) \times 2 + \mathrm{sign}(x_1 - y_1)$	Constant-error
62	Perceived relative argument model	$U(X) = \mathrm{I}(x_1 > y_1)\left(\frac{x_1}{y_1}\right)^{\zeta} + \mathrm{I}(y_1 > x_1)\left(\frac{p_1}{q_1}\right)^{(p_1 + q_1)^{\nu}}$	Logit, Probit or Constant-error

Note. In order to ensure that the KL divergence between two series of model predictions is tractable, choice probabilities $p[X;Y]$ for all models are bounded within the interval [0.001, 0.999]. The following additional functions are used in Supplementary Table 2:

- $\mathrm{sign}(\cdot)$ is a sign function that returns 1 if the argument is positive, -1 if negative and 0 if zero.
- $\mathrm{I}(\cdot)$ is an indicator function that returns 1 if the argument is true, 0 otherwise.

- Power value function: $u(x) = \text{sign}(x) \cdot |x|^{\gamma}$
- Prospect theory value function: $u_{PT}(x) = \text{sign}(x) \cdot \zeta^{\text{I}(x<0)} \cdot |x|^{\gamma}$
- Relative value function for the utility-weighted sampling model: $u_R(x) = \frac{x}{\max\{x_1, y_1\} - \min\{x_2, y_2\}}$
- Tversky and Kahneman's (1992) probability weighting function: $w_{TK}^{+}(p) = \frac{p^{\tau}}{(p^{\tau} + (1-p)^{\tau})^{1/\tau}}$ and $w_{TK}^{-}(p) = \frac{p^{\theta}}{\left(p^{\theta} + (1-p)^{\theta}\right)^{1/\theta}}$
- Karmarkar's (1978) probability weighting function: $w_K(p) = \frac{p^{\tau}}{p^{\tau} + (1-p)^{\tau}}$
- Lattimore et al.'s (1992) probability weighting function: $w_{LBW}^{+}(p) = \frac{\delta p^{\tau}}{\delta p^{\tau} + (1-p)^{\tau}}$ and $w_{LBW}^{-}(p) = \frac{\mu p^{\theta}}{\mu p^{\theta} + (1-p)^{\theta}}$
- Gonzalez and Wu's (1999) probability weighting function: $w_{GW}^{+}(p) = \frac{\delta p^{\tau}}{\delta p^{\tau} + (1-p)^{\tau}}$ and $w_{GW}^{-}(p) = \frac{\mu p^{\theta}}{\mu p^{\theta} + (1-p)^{\theta}}$
- Prelec's (1992) probability weighting function: $w_P^{+}(p) = \exp\{-\beta(-\ln(p)^{\tau})\}$ and $w_P^{-}(p) = \exp\left\{-\alpha\left(-\ln(p)^{\theta}\right)\right\}$
- Venture theory's payoff-dependent probability weighting function (Hogarth and Einhorn 1990): $w_V(p, x) =$

  $\exp\{-b(-\ln(p)^{t})\}$, with $t = \begin{cases} 1 - \tau \cdot \left|\frac{x}{\max\{x_+, y_+\}}\right| & \text{if } x \geq 0 \\ 1 - \theta \cdot \left|\frac{x}{\min\{x_-, y_-\}}\right| & \text{if } x < 0 \end{cases}$ and $b = \begin{cases} 1 + \beta \cdot \left|\frac{x}{\max\{x_+, y_+\}}\right| & \text{if } x \geq 0 \\ 1 - \delta \cdot \left|\frac{x}{\min\{x_-, y_-\}}\right| & \text{if } x < 0 \end{cases}$. $\max\{x_+, y_+\}$ is the largest

  payoff in the design and $\min\{x_-, y_-\}$ is the smallest payoff in the design.
- Regret (or rejoice) function: $\text{R}(d) = \text{I}(d > 0)\kappa \cdot |d|^{\alpha} - (d < 0)\lambda \cdot |d|^{\beta}$

Table A3

Summary of intertemporal decision models.

ID	Model	Category	Authors	Year	Source (Journal or Book)
1	Exponential	Delay discounting	Samuelson	1937	Review of Economic Studies
2	Hyperbolic	Delay discounting	Mazur	1987	The Effect of Delay and Intervening Events on Reinforcement Value
3	Hyperbolic w/ power time (Mazur)	Delay discounting	Mazur	1987	The Effect of Delay and Intervening Events on Reinforcement Value
4	Generalized hyperbolic	Delay discounting	Loewenstein and Prelec	1992	Quarterly Journal of Economics
5	Exponential time	Delay discounting	Roelofsma	1996	Acta Psychologica
6	Quasi-hyperbolic	Delay discounting	Laibson	1997	Quarterly Journal of Economics
7	Hyperbolic w/ power denominator	Delay discounting	Green and Myerson	2004	Psychological Bulletin
8	Hyperbolic w/ power time (Rachlin)	Delay discounting	Rachlin	2006	Journal of the Experimental Analysis of Behavior
9	Constant sensitivity	Delay discounting	Ebert and Prelec	2007	Management Science

10	Double exponential	Delay discounting	McClure et al.	2007	Journal of Neuroscience
11	Fixed cost	Delay discounting	Benhabib et al.	2010	Games and Economic Behavior
12	Generalized hyperbolic w/ increasing elasticity	Delay discounting	Scholten et al.	2014	Cognitive Science
13	Dual systems	Delay discounting	Loewenstein et al.	2015	Decision
14	Interval	Interval discounting	Read	2001	Journal of Risk and Uncertainty
15	Common aspect attenuation	Interval discounting	Green et al.	2005	Journal of Experimental Psychology: Learning, Memory & Cognition
16	Generalized interval	Interval discounting	Scholten and Read	2006	Management Science
17	As-soon-as-possible	Interval discounting	Kable and Glimcher	2010	Journal of Neurophysiology
18	Generalized interval w/ increasing elasticity	Interval discounting	Scholten et al.	2014	Cognitive Science
19	Similarity w/ difference	Time-as-attribute	Leland	2002	Economic Inquiry
20	Similarity w/ ratio	Time-as-attribute	Leland	2002	Economic Inquiry
21	Additive utility	Time-as-attribute	Killeen	2009	Psychological Review

22	Tradeoff model	Time-as-attribute	Scholten and Read	2010	Psychological Review
23	DRIFT	Time-as-attribute	Read et al.	2013	Journal of Experimental Psychology: Learning, Memory & Cognition
24	Attribute-based model w/ power transformations	Time-as-attribute	Dai and Busemeyer	2014	Journal of Experimental Psychology: General
25	Generalized tradeoff model	Time-as-attribute	Scholten et al.	2014	Cognitive Science
26	Intertemporal choice heuristics	Time-as-attribute	Ericson et al.	2015	Psychological Science
27	Proportional difference	Time-as-attribute	Cheng and González-Vallejo	2016	Decision

Table A4.

Functional forms of intertemporal decision models. The notations are designed for choices between $X = (\$x, t)$ *and* $Y = (\$y, s)$*, where* $y > x > 0$*,.* $s > t \geq 0$*.* $U(X)$ *denotes the utility or choice propensity of X and* $U(Y)$ *denotes the utility or choice propensity of Y. For delay discounting models,* $U(Y)$ *is not presented but can be obtained by replacing* x *and* t *in* $U(X)$ *with* y *and* s*. For time-as-attribute models that represent options' advantages on an ordinal scale,* $A(X)$ *denotes the argument for option X and* $A(Y)$ *denotes the argument for option Y. Free parameters are denoted by Greek letters, with corresponding domains and prior distributions shown in Table A5.*

ID	**Model**	**Function**	**Stochastic specification**
1	Exponential	$U(X) = \delta^t u(x)$	Logit, Probit or Constant-error
2	Hyperbolic	$U(X) = \frac{u(x)}{1 + \alpha t}$	Logit, Probit or Constant-error
3	Hyperbolic w/ power time (Mazur)	$U(X) = \frac{u(x)}{1 + \alpha t^{\beta}}$	Logit, Probit or Constant-error
4	Generalized hyperbolic	$U(X) = \frac{u(x)}{(1 + \alpha t)^{\beta/\alpha}}$	Logit, Probit or Constant-error

5	Exponential time[5]	$U(X) = e^{-\beta \cdot \frac{1}{\alpha}\log(1+\alpha t)} u(x)$	Logit, Probit or Constant-error
6	Quasi-hyperbolic	$U(X) = \begin{cases} \delta^t u(x) = u(x), & \text{when } t = 0 \\ \mu\delta^t u(x), & \text{when } t > 0 \end{cases}$	Logit, Probit or Constant-error
7	Hyperbolic w/ power denominator	$U(X) = \dfrac{u(x)}{(1+\alpha t)^{\beta}}$	Logit, Probit or Constant-error
8	Hyperbolic w/ power time (Rachlin)	$U(X) = \dfrac{u(x)}{1+\alpha t^{\beta}}$	Logit, Probit or Constant-error
9	Constant sensitivity	$U(X) = e^{-(\beta t)^{\alpha}} u(x)$	Logit, Probit or Constant-error
10	Double exponential	$U(X) = (\omega\delta^t + (1-\omega)\tau^t)u(x)$	Logit, Probit or Constant-error
11	Fixed cost	$U(X) = \begin{cases} \delta^t u(x) = u(x), & \text{when } t = 0 \\ \delta^t u(x-\mu x), & \text{when } t > 0 \end{cases}$	Logit, Probit or Constant-error
12	Generalized hyperbolic w/ increasing elasticity	$U(X) = \dfrac{u_{SRS}(x)}{(1+\alpha t)^{\beta/\alpha}}$	Logit, Probit or Constant-error
13	Dual systems	$U(X) = (\omega\delta^t + (1-\omega)\tau^t)u(x)$	Logit, Probit or Constant-error
14	Interval	$U(X) = u(x)\delta^{t^{\tau}}$ $U(Y) = u(y)\delta^{t^{\tau}+(s-t)^{\tau}}$	Logit, Probit or Constant-error

[5] The original exponential time discounting model uses log(t) to transform delay t. Since this function cannot adequately accommodate $t = 0$, we have replaced it with $\cdot \frac{1}{\alpha}\log(1+\alpha t)$ in line with the specification of Scholten et al. (2014). This revision has made this model mathematically equivalent to Loewenstein and Prelec's (1992) generalized hyperbolic discounting model because $e^{-\beta \cdot \frac{1}{\alpha}\log(1+\alpha t)} = \frac{1}{(1+\alpha t)^{\beta/\alpha}}$.

15	Common aspect attenuation	$U(X) = \dfrac{u(x)}{1+\alpha\mu t}$ $U(Y) = \dfrac{u(y)}{1+\alpha\big(\mu t + (s-t)\big)}$	Logit, Probit or Constant-error
16	Generalized interval	$U(X) = \dfrac{u(x)}{\left(1+\alpha t^{\tau\zeta}\right)^{\beta/\alpha}}$ $U(Y) = \dfrac{u(y)}{\left(\left(1+\alpha t^{\tau\zeta}\right)(1+\alpha(s^{\tau}-t^{\tau})^{\zeta})\right)^{\beta/\alpha}}$	Logit, Probit or Constant-error
17	As-soon-as-possible	$U(X) = \dfrac{u(x)}{1+\alpha t}$ $U(Y) = \dfrac{u(y)}{(1+\alpha t)\cdot\big(1+\alpha(s-t)\big)}$	Logit, Probit or Constant-error
18	Generalized interval w/ increasing elasticity	$U(X) = \dfrac{u_{SRS}(x)}{\left(1+\alpha t^{\tau\zeta}\right)^{\beta/\alpha}}$ $U(Y) = \dfrac{u_{SRS}(y)}{\left(\left(1+\alpha t^{\tau\zeta}\right)(1+\alpha(s^{\tau}-t^{\tau})^{\zeta})\right)^{\beta/\alpha}}$	Logit, Probit or Constant-error
19	Similarity w/ difference	$A(X) = \mathrm{I}(s-t>\beta)$ $A(Y) = \mathrm{I}(y-x>\alpha)$	Constant-error
20	Similarity w/ ratio	$A(X) = \mathrm{I}\left(\frac{t}{s}<\tau\right)$ $A(Y) = \mathrm{I}\left(\frac{x}{y}<\gamma\right)$	Constant-error

21	Additive utility	$U(X) = x^{\gamma} - \kappa t^{\tau}$ $U(Y) = y^{\gamma} - \kappa s^{\tau}$	Logit, Probit or Constant-error
22	Tradeoff model	$U(X) = \frac{\kappa}{\beta}(\log(1+\beta s) - \log(1+\beta t))$ $\mathrm{U}(Y) = \frac{1}{\alpha}(\log(1+\alpha y) - \log(1+\alpha x))$	Logit, Probit or Constant-error
23	DRIFT	$U(X) = \kappa(s-t)$ $U(Y) = \tau\left(\left(\frac{y}{x}\right)^{\frac{1}{s-t}} - 1\right) + (1-\tau)\gamma\frac{y-x}{x} + (1-\tau)(1-\gamma)(y-x)$	Logit, Probit or Constant-error
24	Attribute-based model w/ power transformations	$U(X) = \kappa(s^{\tau} - t^{\tau})$ $U(Y) = y^{\gamma} - x^{\gamma}$	Logit, Probit or Constant-error
25	Generalized tradeoff model	$\mathrm{U}(X) = \frac{\kappa}{\lambda}\log\left(1 + \lambda\left(\frac{\frac{1}{\beta}(\log(1+\beta s) - \log(1+\beta t))}{\zeta}\right)^{\zeta}\right)$ $\mathrm{U}(Y) = \frac{1}{\alpha}(\log(1+\alpha y) - \log(1+\alpha x))$	Logit, Probit or Constant-error
26	Intertemporal choice heuristics	$U(X) = \kappa\left(\tau(s-t) + (1-\tau)\frac{2(s-t)}{s+t}\right)$ $U(Y) = \gamma(y-x) + (1-\gamma)\frac{2(y-x)}{y+x}$	Logit, Probit or Constant-error
27	Proportional difference	$U(X) = \frac{s-t}{s} + \eta$ $U(Y) = \frac{y-x}{y}$	Logit, Probit or Constant-error

Note. In order to ensure that the KL divergence between two series of model predictions is tractable, choice probabilities $p[X; Y]$ for all models are bounded within the interval [0.001, 0.999]. The following additional functions are used in Supplementary Table 4:

- I(·) is an indicator function that returns 1 if the argument is true, and 0 otherwise.
- Power value function: $u(x) = x^{\gamma}$.
- Increasingly elastic value function as in Scholten et al. (2014): $u_{SRS}(x) = (1 - \omega)x^{1-\omega} + \lambda\omega x^{\omega}$.

Table A5

Parameter bounds and prior distributions for both risky and intertemporal decision models.

Parameter	**Domain**	**Prior distribution**
$\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\varepsilon}, \boldsymbol{\kappa}, \boldsymbol{\lambda}, \boldsymbol{\rho}, \boldsymbol{\chi}, \boldsymbol{\psi}, \boldsymbol{\upsilon}, (\boldsymbol{\zeta} - \mathbf{1})$	[0, +∞)	Exponential (rate = 1)
$\boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\theta}$	[0, 1]	U(0, 1)
$\boldsymbol{\omega}$	[0.5, 1]	U(0.5, 1)
$\boldsymbol{\eta}$	[-2, 2]	U(-2, 2)
$\boldsymbol{\nu}, \boldsymbol{\phi}$	(-∞, +∞)	N(0, 1)
$(\boldsymbol{\iota} - \mathbf{1})$	Integer, [0, 49]	Binomial (49, 0.5)

Note. $(\zeta - 1)$ has the domain of [0, +∞), meaning that the domain of ζ is [1, +∞). Similarly, $(\iota - 1)$ has the domain of [0, 49] (integer), meaning that the domain of ι is [0, 49] (integer).