Agnostic Utility Function Approximation via Cubic Bezier Splines

Sangil Lee[ab*], Chris M. Glaze[a], Eric T. Bradlow[b], Joseph W. Kable[a]

[a] Department of Psychology, School of Arts & Sciences, University of Pennsylvania

[b] Department of Marketing, Wharton School, University of Pennsylvania

[*]Corresponding Author, sangillee3rd@gmail.com

Agnostic Utility Function Approximation via Cubic Bezier Splines

Utility functions are heavily used for modeling economic decision-making such as intertemporal and risky choice. Researchers have used both parametric and non-parametric approaches to model utility functions. Parametric models are simple and easy to fit, but they often require formal assumptions that can be hard to justify, especially given the myriad of different models that exist with no consensus. Non-parametric approaches require relatively large datasets and specially designed procedures to elicit the utility functions. Here we present a novel method – cubic Bezier splines (CBS) – as an 'agnostic' model that approximates utility functions without rigid formal assumptions. CBS is flexible like non-parametric approaches but does not require large or specially designed datasets. CBS provides smooth functions like parametric models but without many of their assumptions. We demonstrate via simulation and real data that CBS accurately recovers a diverse variety of utility models and shows higher predictive accuracy than alternatives. We also demonstrate that CBS can detect important heterogeneity and consistent granular patterns in choice data that cannot be detected with many parametric models.

Keywords: Flexible modeling, agnostic inference, generalized utility functions

Intertemporal choices (ITCs) and risky choices are one of the most heavily studied decisions across many disciplines. ITCs are decisions regarding outcomes that occur at different times, and risky choices are decisions made regarding probabilistic outcomes. Individuals' preferences in ITC and risky choice have been shown to predict many real world behaviors such as pathological gambling, smoking, health behaviors, susceptibility to mental illness, drug and alcohol abuse, education level and financial status (Alessi & Petry, 2003; Anderson & Mellor, 2008; Brañas, Georgantzis, & Guillen, 2007; Kirby, Petry, & Bickel, 1999; Krain et al., 2008; Lejuez et al., 2005; Lejuez et al., 2003; Schepis et al., 2011; Shamosh & Gray, 2008). Mainly, there are two ways of constructing utility models for ITC and risky choice: parametric and non-parametric. We provide a brief overview of both approaches below and then propose a new method that has the benefits of both but the drawbacks of neither.

Parametric models are the simplest way of describing ITC and risky choice. Researchers can use established parametric models to succinctly describe choice preferences using one or two parameters. For example, in ITC, there is the exponential function (Samuelson, 1937; eq. 1a), where *A* denotes the amount of the delayed reward and *D* denotes the time delay until the receipt of the reward. The free parameter *k* reflects how steeply one discounts delayed amounts. In risky choice, there is the expected utility theory (EUT; eq. 1b), where *A* denotes the amount of the risky reward and *p* denotes the probability of winning the said reward. The free parameter α reflects whether people are risk-seeking ($\alpha > 1$) or risk-averse ($\alpha < 1$). Typically, these models are then placed inside logit or probit forms for choice modeling.

$$(a)\ U_{ITC} = A \cdot e^{-kD}, \qquad (b)\ U_{RC} = A^{\alpha} \cdot p \tag{1}$$

As seen from these two examples, parametric utility models can distill the entire preference into one or two parameters, making them easy to estimate with relatively few choices.

The main drawback of these formal parametric models is that the researcher must assume a specific formula (hence 'formal'). This assumption is often untenable as evidence shows that different people behave according to different utility models (Bruhin, Fehr-Duda, & Epper, 2009)[1]. Furthermore, there is no consensus on what the 'best' model is. Popular models of ITC, in addition to the exponential, include the hyperbolic (Herrnstein, 1981; eq. 2a), the quasi-hyperbolic (Laibson, 1997; eq. 2b), and the generalized hyperbolic function (Green, Fry, & Myerson, 1994; eq. 2c). Doyle's 2010 survey lists 27 different models of ITC (Doyle, 2010).

$$(a)\ U_{ITC} = \frac{A}{1+kD}\ , \qquad (b)\ U_{ITC} = A \cdot \beta e^{-kD}, \qquad (c)\ U_{ITC} = \frac{A}{(1+kD)^s} \tag{2}$$

Popular models of risky choice, in addition to EUT, include the risk-return (Markowitz, 1959; 3a), probability discounting (Rachlin, Raineri, & Cross, 1991; 3b), and prospect theory models[2] (Tversky & Kahneman, 1992; 3c), to name a few.

$$(a)\ U_{RC} = p \cdot A - b \cdot Var, \qquad (b) U_{RC} = \frac{A}{1+h\Theta}, \qquad (c)\ U_{RC} = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{\frac{1}{\gamma}}} \cdot A^\alpha \tag{3}$$

Var (eq. 3a) refers to the variance of the outcome, Θ (eq. 3b) is the odds against the receipt of the reward (i.e. (1-p) / p). Each of these models also have their own variants as well: coefficient of variation (Weber, Shafir, & Blais, 2004; modifies 3a), the generalized hyperbolic probability discounting model (Green, Myerson, & Ostaszewski, 1999; modifies 3b). Prospect theory (eq. 3c) especially has many versions (Stott, 2006) such as the weighting function proposed by Goldstein and Einhorn (Goldstein & Einhorn, 1987; eq. 4a), and by Prelec (1998;eq. 4b).

$$(a)\ U_{RC} = \frac{\delta p^\gamma}{\delta p^\gamma + (1-p)^\gamma} \cdot A^\alpha, \qquad (b)\ U_{RC} = \exp[-\delta(-\ln p)^\gamma] \cdot A^\alpha. \tag{4}$$

One thing is clear: there are too many models for any single study to entertain; and even if one tried to, one would be still susceptible to scholars who may suggest yet another utility function.

An alternative approach is non-parametric. Non-parametric approaches do not assume a formula for the utility function. Nevertheless, their drawbacks discourage their widespread use. One drawback is that the dataset needs to be tailored to the non-parametric approach. For example, an often used non-parametric technique is to precisely measure the utility function at a few points and to interpolate between them (Abdellaoui, 2000; Gonzalez & Wu, 1999; Myerson, Green, & Warusawitharana, 2001; Wakker & Deneffe, 1996). Concordantly, one must measure the utility function at a single point multiple times until it is precisely estimated. Hence, non-parametric approaches cannot be used on datasets where this is not the case. A second drawback is that non-parametric approaches typically require a lot of data. For example, a non-parametric elicitation by Gonzalez & Wu (1999) asked each participant 990 questions. For these two reasons, non-parametric approaches are less popular than parametric models.

For those whose main goal is not in identifying the 'correct' model, these two approaches present a dilemma. One must either assume the form, or painstakingly estimate it with substantial data. To overcome this burden, here we present a novel method with the benefits of both parametric and non-parametric approaches. First, like non-parametric approaches, our method does not assume a particular model and can approximate a wide variety of functions. This frees researchers from model selection as our method can be a one-size-fits-all approach. Second, like parametric models, our method can work on small datasets and provides inherently smooth functions. With these combined benefits, we show that our method has higher predictive power and can discover consistent individual behavioral patterns that current parametric models miss.

The statistical motivation for our method is straightforward. Many utility models share a common form of $U = f(A) \cdot g(X)$, where A would be the monetary amount and X would be the delay (D) in ITC or the probability (p) in risky choice. Even non-parametric methods that do not

assume formulas for $f(A)$ and $g(X)$ assume this multiplicative relationship between them (Abdellaoui, 2000; Gonzalez & Wu, 1999; Myerson et al., 2001; Wakker & Deneffe, 1996). Some examples of these functional forms are shown in Figure 1.

[INSERT FIGURE 1 HERE]

Given the flexible shapes for the extant literature shown in Figure 1, we sought a method that could nest the variety of shapes without rigid formulaic assumptions: an agnostic method. But we also knew that the current non-parametric methods require substantial amount of data due to their flexibility. Hence, to combat the over-flexibility, we wanted to use two implicit constraints built into parametric models: smoothness and monotonicity. Smoothness constrains the function from making abrupt changes while monotonicity constrains the direction of the change. These constraints reduce the potential space of functions and fight over-flexibility.

There were several possible candidates for such agnostic models[3]. Polynomial splines, or more generally, basis-function based models can flexibly, smoothly model functions. However, they are difficult to constrain monotonically because this requires constraining its derivative, which evolves as the order of polynomial spline increases. Kernel-based or data-smoothing methods are also hard to constrain for monotonicity since they lack a formulaic definition. Piecewise linear splines could easily be constrained but are not inherently smooth. Modern machine learning methods such as random forests or neural networks do not yield interpretable structure in the form of $U = f(A) \cdot g(X)$ as they deal with a generalized function such as $U = f(A,X)$. We found that cubic Bezier splines (De Casteljau, 1963) can provide a solution that satisfies these conditions as it is smooth, flexible and easy to constrain as we describe next.

PROPERTIES OF CUBIC BEZIER SPLINES

Bezier splines, (especially cubic Bezier splines) are used frequently in computer graphics, fonts, animations, and much more. A Cubic Bezier Spline (CBS) is defined by four points known as control points (Fig. 2). The spline starts from the first point and ends at the fourth point (P0 and P3 in Fig. 2) but it usually does not pass through the middle two points (P1 and P2 in Fig. 2). For clarity, we refer to the first and the fourth control points as anchor points in this manuscript.

[INSERT FIGURE 2 HERE]

A CBS has two useful properties. The first is the convex hull property: the spline is always inside the polygon created by the control points (e.g., Fig. 2, A ~ F). This property guarantees that the spline is well contained where the control points are (cf., polynomials splines can behave quite erratically from small changes in parameters). The second property is that the spline's local derivative at an anchor point equals the slope of the line connecting the anchor point with its next control point (i.e., $\overline{P_0P_1}$). Using this property, multiple pieces of CBS can be smoothly joined by ensuring the anchor point and its two control points are on the same line. This is shown in Figure 2G, with the adjoining point being P_3.

The flexibility of CBS comes from its mathematical form. The *x* and *y*-coordinates of the spline are controlled independently by two separate cubic functions (eq. 5, 6)[4].

$$x = h(t) = (1-t)^3P_{0x} + 3(1-t)^2tP_{1x} + 3(1-t)t^2P_{2x} + t^3P_{3x}, \qquad 0 \le t \le 1 \qquad (5)$$

$$y = g(t) = (1-t)^3P_{0y} + 3(1-t)^2tP_{1y} + 3(1-t)t^2P_{2y} + t^3P_{3y}, \qquad 0 \le t \le 1 \qquad (6)$$

The CBS is drawn by *x* and *y* values that are determined as *t* moves from 0 to 1. When $t = 0$, $x =$

$h(0) = P_{0x}$, and $y = g(0) = P_{0y}$, which are the x and y coordinates of P_0. When $t = 1$, $x = h(1) = P_{3x}$, and $y = g(1) = P_{3y}$. Figure 3 graphically shows this relationship between CBS, $h(t)$, and $g(t)$.

[INSERT FIGURE 3 HERE]

CBS, by itself, is not a function expressing y in terms of x (i.e., $y = f(x)$). Rather, since the form is $y = g(t)$, and $x = h(t)$, we must use $y = g(h^{-1}(x))$. This is done by first finding the real t that satisfies $x = h(t)$ and $0 \leq t \leq 1$. Then t is plugged into $y = g(t)$ to get the y-coordinate. Computing a k-piece CBS function (denoted as $y = CBS^k(x)$ from here on) is as follows. Let the ith piece ($i = 1,\ldots,k$) of CBS in a k-piece function $CBS^k(x)$ be defined by $x = h_i(t)$, and $y = g_i(t)$. First, for given x, find i such that $h_i(0) \leq x \leq h_i(1)$, and then find real t^* such that $x = h_i(t^*)$, and $0 \leq t^* \leq 1$. Then, $y = CBS^k(x) = g_i(t^*)$. The first step identifies the piece whose domain contains x and the second step finds the corresponding t^*, which is then used to calculate y.

Because CBS is a piecewise function, the mathematical constraints for CBS are simply applied to each piece independently. This makes CBS very easy to constrain regardless of the flexibility required. In CBS, each piece can be independently constrained up to its second derivative. Constraining the sign of the first derivative gives us monotonicity, while constraining the sign of the second derivative gives us convexity or concavity. For CBS to approximate functions of the form $y = f(x)$, we must constrain $x = h(t)$ to be a monotonically increasing function of t. This ensures that the spline does not have multiple y values for a given x (e.g., Fig. 2.e). Next, if needed, we can constrain CBS to be monotonically increasing or decreasing by simply enforcing $g'(t) \geq 0$ or $g'(t) \leq 0$, respectively. If each chain is constrained to be monotonically increasing, the whole CBS function will also be monotonically increasing as well. The necessary conditions and their proofs are shown in the web appendix.

Before we move onto choice data where utility functions are latent, we first show that CBS can closely approximate various utility functions. We drew various utility functions with no noise and evaluated how closely CBS can match them. For the 4 functions in ITC (exponential, hyperbolic, generalized hyperbolic, quasi-hyperbolic) and the 4 functions in risky choice (value function: A^{α}, Tversky-Kahneman, Goldstein-Einhorn, Prelec), we varied each of their parameters in 10 levels. $ln(k)$ was varied from -10 to -1, β was varied from 0.1 to 1. Other parameters (s, α, γ, δ) were varied from 0.2 to 2 in increments of 0.2. Then, for each generated function, we sampled 101 evenly spaced points throughout the domain of the function. Then we fit a CBS function to these 101 points by minimizing mean squared error (MSE). We then examined the mean absolute error (MAE) of the 1-piece and 2-piece CBS fits across different parameters of each function. Most functions' range was [0 1], but that of the value function varied depending on the parameter α. To make MAE comparable, we normalized the range of the value functions to be [0 1]. This was done by simply dividing each value function by 100^{α}.

[INSERT FIGURE 4 HERE]

Figure 4 shows example fits that demonstrate CBS's flexibility. The top row shows that a 1-piece CBS can closely approximate all four ITC functions. We see some approximation errors at the kink of the quasi-hyperbolic function, which is expected because the quasi-hyperbolic function is discontinuous; nevertheless, the approximation errors are quite small. The middle two rows show risky choice functions fit by a 1-piece CBS function. We see that the 1-piece CBS shows difficulty with probability weighting functions with sharp kinks near 0. However, these approximation errors are significantly reduced with a 2-piece CBS function (last row).

[INSERT FIGURE 5 HERE]

Across all functions, the MAEs were impressively low (Fig. 5); MAEs were lower than 0.08 for 1-piece CBS and lower than 0.008 for 2-piece CBS. For ITC functions, we see that MAE increases slightly for high discount rates ($ln(k) \sim [-4\ -1]$) because high discount rates result in sharp kinks near 0. We also see that the quasi-hyperbolic shows higher MAEs with smaller β, because a smaller β means greater discontinuity at 0. In risky choice probability weighting functions, we see that MAE increases for very small γ, which is, again, because low γ results in a sharp kink near 0. However, 2-piece CBS provides excellent approximations even in those cases.

Now, can CBS also approximate latent utility functions? In the next two sections, we address this question in ITC and risky choice via simulated function recovery and real data cross-validation. We also use CBS to yield new insights into choice behavior. In ITC, we show there is a sizable group of participants whose behavioral patterns are not consistent with any of the considered parametric models. In risky choice, we use CBS to obtain a granular measure of risk-aversion that shows individual-specific patterns of risky preference at various probabilities.

INTERTEMPORAL CHOICE

In this section, we evaluate CBS's performance in agnostic utility function approximation in ITC. Often-used ITC models are in the same general form[5]: $U = A \cdot g(D)$. These include the aforementioned exponential, hyperbolic, generalized hyperbolic, and the quasi-hyperbolic (Green et al., 1994; Herrnstein, 1981; Laibson, 1997; Samuelson, 1937). Here, we use CBS to estimate the function $g(D)$ without making formal parametric assumptions.

In the first half of this section, we use simulation to assess how well CBS can recover the choice-generating utility functions. After simulating choices from known utility functions, we measure the MAE of the fitted CBS functions. In the second half, we use a real ITC dataset to 1) examine the cross-validation performance of CBS and 2) use CBS to detect participants that cannot be described well by the existing parametric models. There are several reasons why formal parametric models may not be able to describe certain behavioral patterns. Consider a person's discount rate over time. This is the rate at which the value of the delayed reward depreciates. All four parametric models we consider here have either a constant discount rate (exponential; value of delayed reward is reduced by a fixed percentage every unit time) or decreasing discount rates (hyperbolics) over time. However, a person may have an increasing discount rate. If a person has a threshold beyond which one would not wait for any improvement in outcome, that person's discount rate will sharply increase at the given delay (e.g., I need class material by tomorrow, but I don't need it after that).

ITC Simulation Analyses

Methods. We simulated choices from various utility functions and used CBS to recover the utility functions. We chose 20 utility functions: 4 from each of exponential, hyperbolic, generalized hyperbolic, quasi-hyperbolic, and non-parametric functions ($ln(k)$ = -2, -4, -6, -8 for exponential and hyperbolic; $ln(k)$ = -4, -7 and s = 0.5, 2 for generalized hyperbolic; $ln(k)$ = -4, -7 and β = 0.4, 0.7 for quasi-hyperbolic). The 4 non-parametric functions reflected a variety of general patterns: 1) an amount heuristic (all delayed options are worth 50%), 2) a delay heuristic (no waiting beyond 100 days), 3) a linearly declining function ($g(D) = 1 - D / 180$), and 4) a concave function ($g(D) = [(D/180)^2 - 1]^{0.5}$).

Each simulated choice was between a smaller, immediate monetary amount of $20 (fixed

across all trials), and a larger, delayed monetary amount whose delay and amount varied for each trial. These amounts and delays were sampled uniformly from a range of delays (1~180 days), and a range of immediate to delayed amount ratio (0 ~ 1; e.g., ratio of 0.5 means the immediate amount is half the delayed amount). Dataset sizes ranged from 7^2 to 20^2 choices based on how finely we sampled the range of delay and amount. The difference in utilities of the two options (i.e., immediate \$20 and delayed), were used in a logit model (eq. 7).

$$p(delayed\ choice) = \left[1 + \exp\left(-\phi(A \cdot g(D) - 20)\right)\right]^{-1} \quad (7)$$

For each of the 20 functions x 14 dataset size conditions, we simulated 1000 datasets. The scaling parameter ϕ was fixed at 1, as it is not a variable of interest in our study.

Each of the simulated datasets was then fitted with 1-piece CBS and 2-piece CBS using the same logit model specification (eq. 7; $g(D) = CBS^k(D)$). The CBS functions were constrained to be monotonically decreasing, and $CBS^k(0) = 1$. After these constraints, 1-piece CBS required 5 parameters, and 2-piece CBS required 10 parameters. The scaling parameter ϕ was also estimated to emulate model fitting in real data where we do not know the scaling parameter. We measured the mean absolute error (MAE) between the fitted CBS functions and the true simulating functions to assess CBS's recovery of true functions. Since the error is measured relative to the true function, it is not a given that more flexible models will have lower MAEs as it may overfit the choice noise instead of the true function.

Results. CBS shows excellent recovery of various latent utility functions. Fig. 6 shows the true functions and their average CBS fits from 20^2 choice dataset condition. By examining the average fit, we can assess whether CBS has systematic biases or shortcomings in recovering the functions. In the top panels, we see that both 1-piece CBS and 2-piece CBS show little error

in approximating the parametric ITC functions. As before, we do see that 1-piece CBS has more difficulty with sharp kinks near 0 in quasi-hyperbolic simulations and the steepest exponential functions. However, MAEs are quite small for reasonable delays (MAE < 0.05 for delay > 10). In the bottom panels showing non-parametric simulations, we see that 1-piece CBS has difficulty with delay heuristics while 2-piece CBS fares batter. The MAEs for both 1-piece and 2-piece CBS are less than 0.02 for the other non-parametric functions.

[INSERT FIGURE 6 HERE]

CBS shows little error even at smaller dataset sizes. Table 1 shows the mean MAEs of CBS fits under different simulating functions and dataset sizes. Note that this is not the MAE of the average fit (which is Fig. 6), but rather the average MAE of the fits. The former shows CBS fits after averaging out the choice noise, and the latter shows how much CBS fits are deterred by choice noise. The latter also tells us the expected MAE for CBS fits under different dataset sizes. Concordantly, we see that the mean MAEs decrease as the size of dataset gets larger. Nevertheless, the expected MAEs are very small in all dataset sizes (below 0.04 for most cases except for 1-piece CBS in non-parametric). Generally, we see that 1-piece CBS has lower errors than 2-piece CBS except when the underlying functions are difficult to approximate with 1-piece (e.g., exponential and quasi-hyperbolic with sharp kinks and non-parametric functions).

[INSERT TABLE 1 HERE]

ITC Real Data Analyses

Methods. We obtained real ITC data from Kable et al. (2017). They examined changes in ITC and risky choice after 10 weeks of commercial cognitive training; the original authors reported no significant changes in ITC and risky choice after 10 weeks. 166 participants

completed session 1 and 128 came back for session 2. In each session, they completed ITC and risky choice tasks. We treated the dataset as if there were 294 participants since we wanted to ensure that all participants have a similar number of datapoints, and we did not know a priori whether the behavior was consistent enough across sessions to warrant merging the 2 sessions' data. We only use the ITC dataset here, while the risky choice dataset is used in the next section.

In each session, participants made 120 binary choices in the ITC task. The choices were between a smaller immediate monetary reward that was always $20 today (i.e., the day of the experiment) and a larger later monetary reward (e.g., $Y in X days; X ~ [20 180], Y ~ [22,85]). We only included sessions with at least two or more of each choice types (i.e., choice of the larger delayed or smaller immediate reward), which ruled out 9 sessions and brought the total data to 285 sessions. This was because at least two of each choice type was necessary for leave-one-trial-out cross validation; otherwise the training dataset may have entirely one-sided choices which wouldn't allow the model to be estimated. For cross-session correlation analysis, we only included participants who had both sessions' data (122 participants, 244 sessions of data).

We assessed leave-one-out cross-validation (LOOCV) performances of 8 ITC models. Six of these were those used in our simulation analyses: exponential, hyperbolic, generalized hyperbolic, quasi-hyperbolic, 1-piece CBS, and 2-piece CBS. The other two models were arithmetic discounting (Killeen, 2009) and the fixed cost model (Benhabib, Bisin, & Schotter, 2010), which are models that cannot be expressed as $U = A \cdot g(D)$. LOOCV was performed by training the 8 models with MLE (eq. 7) on $n - 1$ trials and predicting the left-out trial. This was repeated until every trial was predicted. We obtained two LOOCV performance metrics. First, we calculated the accuracy (hit rate) of the predictions based on a threshold of $p(\textit{delayed choice}) = 0.5$. However, simple hit rate has two problems: 1) choices are unbalanced and a bad model

that only predicts one type of choice might also have high accuracy, and 2) it doesn't take into account the predicted choice probabilities. So, we also obtained Tjur's D (Tjur, 2009). Tjur's D (coefficient of discrimination) is the difference of the mean choice probabilities of each choice type. A good model should have high *p*(*delayed choice*) for delayed choices but low *p*(*delayed choice*) for immediate choices. Hence, the difference between the mean of those two choice probabilities is bounded in [0 1] and tells how well the two choice types are discriminated in out-of-sample predictions. Even if two models have the same accuracy, Tjur's D would be higher for models that classify the trials with larger discrimination in choice probabilities. LOOCV accuracies and Tjur's D were compared between models using paired *t*-tests with holm-bonferroni correction for family-wise error rate.

To detect and describe the heterogeneity of ITC utility functions, we calculated the moment-by-moment log discount rates for each individual. We took the estimate of the discounting function $f(D)$ for each individual using the CBS (one-piece or two-piece) with the higher LOOCV accuracy. Then we calculated the log of the daily exponential discount rate, $\ln(-\ln f(D) / D)$, across the range of D (we use the log as we assume discount rates are log-normally distributed). Concordantly, exponential discounting functions have a constant daily discount rate. If one's discounting function is e^{-kD}, $k > 0$, their moment by moment discount rate is $\ln(-\ln(e^{-kD}) / D) = \ln(kD / D) = \ln(k)$. The different hyperbolic functions all have daily discount rates that decrease over time. As mentioned, increases in daily discount rates cannot be accounted for by any of the formal parametric models we considered.

To show that this moment-by-moment measure of discount rate is a person-specific characteristic just like a person's overall level of discounting, we assessed the participants' cross-session consistency between the moment-by-moment discount rates. For the 122 participants

who had both session 1 and session 2, we calculated a vector of ln(k) (the moment-by-moment discount rate) for each of their two sessions and calculated the Spearman correlation coefficient between them. The mean of the 122 correlation coefficients was taken as a summary statistic regarding average cross-session consistency of moment-by-moment discount rate. Its significance was tested by performing permutation tests (random pairing of session 1 and session 2 data to calculate their average cross session correlation) 10,000 times.

Results. The CBS functions showed higher LOOCV performances than all other parametric models considered, both in terms of accuracy and discrimination. The highest average LOOCV accuracy was by 1-piece CBS (mean = 90.45%, std. error = 0.31%), followed by 2-piece CBS (mean = 90.18%, std. error = 0.31%). 1-piece CBS's LOOCV accuracy was significantly higher than the other 7 models (corrected $p \leq .05$). However, the parametric models also had very high LOOCV accuracies (*mean, std. error*): fixed cost (90.17%, 0.31%), quasi-hyperbolic (90.15%, 0.31%), generalized hyperbolic (90.07%, 0.31%), exponential (89.21%, 0.31%), arithmetic (89.15%, 0.31%), hyperbolic (89.08%, 0.32%). To better examine the model's predictive powers, we measured each model's discriminating power in LOOCV predictions with Tjur's D (Fig. 7). Here, 2-piece CBS had the highest LOOCV Tjur's D (mean 65.06%, std. error 1.04%) followed by 1-piece CBS (mean 64.30%, std. error 1.04%). 2-piece CBS's Tjur's D was significantly higher than the other 7 models (corrected $p \leq .001$). 1-piece CBS's Tjur's D was significantly higher than the exponential, hyperbolic, generalized hyperbolic, and arithmetic models (corrected $p \leq .005$), but was not higher than the others (corrected $p \geq .05$). While parametric models give high predictive accuracy (although significantly lower than that of CBS), CBS has significantly higher discriminating power.

[INSERT FIGURE 7 HERE]

To provide further insight, we show eight example participants' data and their model fits in Figure 8. Both the fit and the choices are shown in relative amounts. For example, a choice of \$20 vs. \$40 in 6 days is essentially asking if $f(D = 6)$ is greater or less than 0.5, which is the relative amount 20/40. By plotting each question in terms of relative amount and delay, we can see whether the fitted function (drawn in solid black line) is appropriately dividing the delayed choices (shown in circles) and the immediate choices (shown in Xs). Panel A shows 4 participants' data whose highest LOOCV accuracies came from extant parametric models. The top row shows the parametric model fits while the middle row shows the CBS model fits. In these cases, participants' choices were well aligned with known parametric models and CBS shows good approximations of them. Panel B, on the other hand, shows 4 participants' data whose highest LOOCV accuracies came from CBS. We can see in the top row that even the best extant parametric models are unable to separate the choices well. In contrast, CBS (middle row) fits a rather unconventional, but flexible monotonic function that separates the two choice types.

[INSERT FIGURE 8 HERE]

Figure 8 also shows the moment-by-moment log discount rates calculated from CBS fits for the eight example participants. Participants who were best fit by extant parametric models showed constant (in the case of exponential) or decreasing log discount rates over time. On the other hand, participants who were best fit by CBS models showed increasing log discount rates over time. Such increasing discount rate functions cannot be explained by the extant parametric models as they all exhibit non-increasing discount rates over time (examples shown in Fig. 9A).

[INSERT FIGURE 9 HERE]

When grouped according to the model with highest LOOCV accuracy, there was a clear

trend that shows a distinct group of participants who had increasing discount rates over time (Fig. 9B). As theoretically expected (panel A), participants who are best fit by exponential models show a relatively constant moment-by-moment discount rate throughout their domain. Participants who are best fit by other extant parametric models show declining moment-by-moment discount rates as well. These results show that CBS has successfully captured these participants' characteristic discounting function. On the other hand, participants who were best fit by CBS functions show increasing discount rates over time.

To ensure that this increasing discount rate was a stable person-specific characteristic, we measured each participants' Spearman correlation across the two sessions' moment-by-moment discount rate. The mean cross-session correlation coefficient was 0.34 (median = 0.60, interquartile range = [-0.24, 0.97]). Using permutation tests, we generated a null distribution by randomly correlating session 1 and session 2 data. Only three of the 10,000 permutations resulted in a mean cross-session correlation of 0.34 or higher, which puts the *p*-value well below 0.001. This result shows that there are consistent individual differences in the shape of the discounting function that go beyond previously studied differences in the overall level of discounting.

ITC Discussion

In this section, we examined how CBS can be used to model ITC data. We showed via simulation that CBS can approximate many established parametric models as well as 'unconventional behaviors' at various dataset sizes. In a real choice dataset, we showed that CBS has higher LOOCV accuracy compared to all other tested parametric models. We also demonstrated that it can detect heterogeneous utility functions that simply cannot be captured by extant parametric models. Thus, we have shown that in ITC, CBS has not only the benefits of a non-parametric approach by being able to fit a broad class of utility functions, but also the

benefits of extant parametric models that do not require large or specially structured datasets. If one wants to obtain an overall measure of impatience (e.g., hyperbolic k), one can use the area under the curve (AUC) as an agnostic measure of the discount rate.

From a substantive point, by using CBS, we showed that there is a group of participants who displayed increasing daily discount rates as opposed to the usual assumption of constant or decreasing discount rates. This could occur due to a hard 'stopping rule' that makes participants simply unwilling to wait beyond a certain delay. For example, participants who are expecting a significant event in the future that may prohibit one from receiving delayed rewards may exhibit this behavior (e.g., death, moving to a new place). In such circumstances, simply increasing the magnitude of the delayed reward will not have much effect since it is already heavily depreciated. Future studies may examine how often participants exhibit increasing discount rates and what the specific psychological causes of this behavior might be.

RISKY CHOICE

In this section, we evaluate CBS's performance in agnostic utility function approximation in risky choice. In risky choice, there are three different popular forms of risky choice utility functions. First is the prospect theory form: $U = f(A) \cdot g(p)$. In prospect theory, a gamble of winning a large amount (A) with probability p and a smaller amount (SA) with probability $1\text{-}p$ has utility of $U = g(p) \cdot f(A) + (1 - g(p)) \cdot f(SA)$. When the smaller amount is zero, it becomes a simple gamble, and $U = f(A) \cdot g(p)$. Stott's 2006 review of prospect theory forms includes 7 different forms for $f(A)$ and 7 different forms for $g(p)$. Popular ones include the power function

$f(A) = A^{\alpha}$, and the weighting function proposed by EUT, Tversky and Kahneman (1992), Goldstein and Einhorn (1987), and Prelec (1998). The second form of risky choice utility is the probability discounting form: $U = A \cdot g(\theta)$. The discounting function is written in terms of the odds against winning the gamble $\theta = (1 - p) / p$. Like in the ITC, probability discounting has been modeled in a hyperbolic (Rachlin et al., 1991) or a generalized hyperbolic form (Green et al., 1999). Lastly, there is the mean-variance form: $U = EV - b \cdot Var$, where EV is the expected value of the gamble and *Var* is the variance of the gamble (or an alternative measure of variance; e.g., Weber et al., 2004). The mean-variance form does not have a clear latent function that can be estimated, so we do not consider it here.

In this paper, we focus on the discounting form for three compelling reasons. First, prospect theory forms can be analytically converted into discounting forms. Consider a simple gamble and its certainty equivalent: $CE^{\alpha} = A^{\alpha} \cdot g(p)$. We can de-exponentiate and then re-parameterize to achieve the following discounted utility form of risky choice:

$$CE = A \cdot g(\theta + 1)^{-1/\alpha} \tag{8}$$

Some example graphs of this conversion are shown in Figure 10. A small caveat is that this conversion only applies to simple gambles where one of the outcomes is zero. However, many datasets use simple gambles since gambles with two or more non-zero outcomes can be cognitively demanding.

[INSERT FIGURE 10 HERE]

The second reason is that the discounting form is easier to interpret. Because prospect theory posits utility as a product of two functions, neither one of them is solely responsible for the participant's risk preference. Consequently, the prospect theory form does not immediately

illuminate whether one is risk-averse or the degree of the risk-aversion. In the discounting form, however, one can immediately identify the degree of risk-aversion and even identify the range of probabilities at which the person is risk-averse and risk-seeking. This is because there is a discounting curve that corresponds to expected value (the middle graph of EUT in Figure 10). If one's utility function is above this curve, one is risk-seeking, and if otherwise, risk-averse.

The third reason is that the prospect theory form is much harder to identify than the discounting form. This is again because prospect theory is a product of two separate functions; if one's $f(A)$ is generally high, one can lower the $g(p)$ to compensate to achieve the same utility. One can also see this effect from the example functions in Figure 10. In the EUT model, we can see that the value function controls the steepness of the discounting curve. But, we also see that the probability weighting functions also control the steepness of the discounting curve. Fox and Poldrack (2009) noted this difficulty and suggested that researchers take caution. Bruhin et al. (2009) have noted in their paper that "fitting a (prospect theory) model for each individual is … frequently impossible and often not desirable in the first place." We also provide our own simulation results in our web appendix that clearly shows the confound between $f(A)$ and $g(p)$.

Given these three reasons, we use CBS to model probability discounting functions. In the first half of the section, we use simulation to show how well CBS can recover various utility functions especially those in prospect theory forms. In the second half, we use real data to compare CBS's cross-validation performance against other models. Also, we show how CBS can be used to obtain a granular measure of risk-aversion. Typically, previous studies examining risk-aversion have simply obtained one parameter estimate that measures the overall degree of risk-aversion (e.g., parameter α in EUT). Here we show how to use CBS to obtain a richer measure quantifying the pattern of risk-aversion across probabilities.

Risky Choice Simulation Analyses

Methods. We generated choice datasets using prospect theory models in original form[6], fitted CBS to the choices, and compared the CBS fits against the analytically converted prospect theory forms (eq. 8). We chose 28 utility functions to generate choices: 4 expected utility theory models, 8 Tversky-Kahneman models, 8 Goldstein-Einhorn models, and 8 Prelec models (α = 0.5, 0.75, 1, 1.5 for EUT; α = 0.7, 1.3 and γ = 0.25, 0.5, 1, 1.5 for Tversky-Kahneman; α = 0.7, 1.3, γ = 0.5, 2, and δ = 0.5, 2 for both Goldstein-Einhorn and Prelec). As in the ITC simulations, each simulated choice was between a fixed smaller, certain monetary amount of $20 and a larger, risky monetary amount whose probability and amount varied from trial to trial. Each dataset was created by uniformly sampling the range of odds against between 0 and 4, and the range of ratios of certain to risky amount (0 ~ 1). The dataset size ranged from 7^2 choices to 20^2 choices depending on how finely we sampled the range of odds against and the range of ratios.

For each of the 28 simulating functions and 14 dataset sizes, we calculated the utilities of the two options (i.e., certain and risky) and the difference between these two utilities were used in a logistic model to calculate choice probabilities (eq. 9; ϕ = 1).

$$p(risky\ option) = \left[1 + \exp\left(-\phi\left(U_{risky} - U_{certain}\right)\right)\right]^{-1} \quad (9)$$

For each of the 28 functions x 14 dataset size conditions, we simulated 1000 datasets. Each of the simulated datasets was then fitted with CBS functions using the same logistic model (eq. 9): $U = A \cdot CBS^k(D)$, k = 1, 2. The CBS functions were constrained to be monotonically decreasing, and $CBS^k(0) = 1$. The scaling parameter ϕ was also simultaneously estimated.

Results. CBS fits show excellent approximations to prospect theory models as shown by their small errors from the analytically converted discounting forms of the four prospect theory

models. Fig. 11 shows the simulating functions and their average CBS fits under the largest dataset condition (400 choices per dataset). We see that both 1-piece CBS and 2-piece CBS show very little error in approximating the underlying functions. As before, we do see that 1-piece CBS functions have more difficulty dealing with sharp kinks near 0 as can be seen in some Tversky-Kahneman functions. Nevertheless, the errors of these fits were only high near 0.

[INSERT FIGURE 11 HERE]

CBS's recovery of various utility functions is also excellent even at low dataset sizes. Table 2 shows the MAE of CBS fits under different simulation settings for dataset size and true underlying function. Throughout all simulations the average MAEs are around or less than 0.06 and decreases as dataset size increases. In most simulations, we see that 1-piece CBS has lower average MAEs than 2-piece CBS. Of course, in Tversky-Kahneman simulations where 1-piece CBS has systematic difficulties, we see that 2-piece CBS starts to show lower MAEs than 1-piece CBS around 169 choices.

[INSERT TABLE 2 HERE]

Risky Choice Real Data Analyses

Methods. The real data comes from Kable et al. (2017) in a task structure similar to ITC. Each participant made 120 binary choices per session. The choices were between a larger risky monetary reward (e.g., $40 with 50% chance) and a smaller certain monetary reward that was always $20. The risky reward amount ranged from $21 to $85. Out of the 294 total sessions, we only included sessions that had at least two of each choice types (i.e., choice of the larger risky or smaller certain reward; two of each is required for LOOCV). This excluded 4 sessions, bringing the total to 290 sessions. In LOOCV analysis, we treated the dataset as if there were 290

participants since we wanted to ensure that all participants have similar number of datapoints and also since we didn't know a priori whether participants' behavior was consistent enough across sessions to warrant merging the 2 sessions' data. For the cross-session consistency analysis of granular risk-aversion, we only used participants who had both sessions of data (125 participants, 250 sessions).

We performed an extensive leave-one-out cross-validation (LOOCV) exercise using various models[7]. We entertained a total of 11 models: EUT, Tversky-Kahneman, Goldstein-Einhorn, Prelec 1 parameter, Prelec 2 parameter, hyperbolic, generalized hyperbolic, original mean-variance model, Weber's modification to mean-variance model, and two CBS models (1-piece and 2-piece CBS). For each of the 120 trials, we trained the model on 119 trials by using MLE with the logit model specified by eq. 9, and then used the trained model to predict the left out trial. This process was repeated 120 times. Then we calculated the LOOCV prediction accuracies of each model based on the threshold of $p(\textit{risky option}) = 0.5$. We also obtained each model's Tjur's D by subtracting the mean $p(\textit{risky option})$ of certain choice trials from the mean $p(\textit{risky option})$ of risky choice trials. LOOCV accuracies and Tjur's D were compared between models using paired *t*-tests with holm-bonferroni correction for family-wise error rate.

We also used CBS to obtain a more granular pattern of risk aversion by calculating the moment-by-moment risk aversion for each session's data. We took the estimate of the discount function $f(\theta)$ for each session from the CBS (one-piece or two-piece) with the higher LOOCV accuracy. Then, we calculated $\log(f(\theta)^{-1} - 1) - \log(\theta)$ across the range of θ to yield a granular measure of risk aversion. A person who chooses the gamble with the higher expected value would have utility function of $U = A \cdot p = A \cdot (\theta + 1)^{-1}$, which is a hyperbolic discounting function with discount rate of 1. Hence, if one calculates the moment-by-moment hyperbolic

discounting factor h from $(h\theta + 1)^{-1}$, a person is risk-averse if h is greater than 1 and risk-seeking if h is smaller than 1. In logarithmic space, a person is either risk-averse or risk-seeking depending on whether $\log(h)$ is greater than 0 or less than 0, respectively.

To demonstrate that this granular pattern of risk aversion is a person-specific characteristic just like a person's overall level of risk-aversion, we assessed the participants' cross-session consistency between the granular risk aversions. For the 125 participants who had both session 1 and session 2, we calculated a vector of $\log(h)$ (the moment-by-moment risk aversion measure) for each of their two sessions and calculated the Spearman correlation coefficient between them. The mean of the 125 correlation coefficients was taken as a summary statistic regarding average cross-session consistency of our granular measure of risk aversion. Its significance was tested by performing permutation tests (random pairing of session 1 and session 2 data to calculate their average cross session correlation) 10,000 times.

Results. The CBS functions showed great performance in LOOCV prediction both in accuracy and discrimination. 2-piece CBS had the highest LOOCV accuracy followed by 1-piece CBS (2-piece: mean 87.69%, std. error 0.37%; 1-piece: mean 85.55%, std. error 0.38%). 2-piece CBS's LOOCV accuracy was significantly higher than the other 10 models (corrected $p < .001$) and 1-piece CBS's LOOCV accuracy was significantly higher than 7 other models (corrected $p < .001$) but not higher than Goldstein-Einhorn and 2-parameter Prelec functions (corrected $p > .05$). The LOOCV accuracies of the parametric models are as follows in decreasing order (mean, std. error): Goldstein-Einhorn (85.47%, 0.45%), 2-parameter Prelec (85.21%, 0.44%), 1-parameter Prelec (84.59%, 0.46%), Tversky-Kahneman (82.79%, 0.42%), mean-variance (81.36%, 0.42%), generalized hyperbolic (80.81%, 0.38%), EUT (80.37%, 0.39%), hyperbolic (79.49%, 0.41%), and Weber's coefficient of variation (79.43%, 0.42%). CBS also had great

LOOCV discrimination as measured by Tjur's D (Fig. 12). 2-piece CBS had the highest mean LOOCV Tjur's D (mean 64.95%, std. error 1.17%), which was significantly higher than the other 10 models (corrected $p < .001$). 1-piece CBS's Tjur's D was slightly lower than that of the 2-parameter Prelec and Goldstein-Einhorn models but was significantly higher than the other 7 models (corrected $p < .002$).

[INSERT FIGURE 12 HERE]

We then visualized and compared the fits of CBS and other extant parametric models. Figure 13 shows 8 example participants' choice data along with their best parametric model fits and their CBS fits. As in ITC, both the fitted curve and the choices are plotted in relative amounts, which is obtained by dividing the certain amount ($20) by the risky amount. Panel A of Fig. 13 shows 4 participants' data whose highest LOOCV accuracies came from extant models. The top row shows the discounting form of the extant model fits while the bottom row shows the CBS model fits. In these cases, participants' choices were well aligned with known parametric models and CBS shows good approximations of them. Panel B of Fig. 13 shows another 4 participants' data whose highest LOOCV accuracies came from CBS. We can see in the top row that the best extant parametric models are unable to separate the different choices well. On the other hand, the bottom row shows that CBS fits a rather unconventional, but monotonic function that separates the two choice types.

[INSERT FIGURE 13 HERE]

Participants' granular measures of risk-aversion calculated from CBS were surprisingly consistent across the two sessions which were 10 weeks apart. Figure 14 shows three example participants' data along with their moment-by-moment risk aversion measure $\log(h)$. When

log(*h*) is above 0, the participant is showing risk-averse behavior (and more so the farther away from 0), and when log(*h*) is below 0, the participant is risk-seeking. As seen from Figure 14, CBS allows us to measure a rich granularity of risk aversion that traditional models cannot. Participant 1 is risk-seeking at low odds (high probabilities) but becomes risk-averse from odds of 1 (50% probability). Participant 2 is risk-seeking at two points and also risk-averse at two points. Participant 3 is risk-averse at both ends of probability but risk-seeking in the middle. These granular patterns of risk-aversion are consistent across sessions. To quantify this stability, we measured each participants' Spearman correlation of granular risk aversion across sessions. The mean cross-session correlation was 0.46, which is a sizable correlation (median = 0.68, interquartile range = [0.06, 0.93]). Using permutation tests, we generated a null distribution by randomly correlating session 1 and session 2 data; none of the 10,000 permutations resulted in mean cross-session correlation of 0.46 or higher. This shows that there are consistent granular patterns of risk aversion that are not captured by the overall level of risk-aversion.

[INSERT FIGURE 14 HERE]

Risky Choice Discussion

In this section, we examined how CBS can be used to model risky choice. We showed via simulation that CBS can approximate both discounting models and prospect theory form models well via the probability discounting framework. In a real choice dataset, we showed that CBS had highest LOOCV accuracy and discrimination out of all models tested. Furthermore, CBS revealed a granular pattern of risk aversion that varied across participants but was consistent within a participant across time.

Much of the discussion we provided in the ITC section also applies here. First, we

recommend that researchers use cross validation to find the best number of CBS pieces for each participant. Our LOOCV results consistently prefers 2-piece CBS which may be because risky choice models tend to have more complex forms than the ITC models, thereby requiring additional flexibility. However, this could vary depending on the choice dataset size. Second, one should be able to obtain a general level of risk aversion by calculating the area under the curve for the fitted function, while being mindful of the sampled datapoints and extrapolation. That being said, our results indicate that the granular pattern of risk-aversion is consistent within a person and future studies may hence investigate if the granular pattern is predictive of other behaviors above and beyond the overall average level of risk-aversion.

In risky choice, researchers have more freedom in deciding how to fit CBS. Here we used CBS to fit discounting in the space of the odds against, $(1 - p) / p$. However, one can fit a CBS discounting function in alternate spaces such as $1 - p$ or $-\log(p)$ which are all non-linear, monotonic transformations of the odds $(1 - p) / p$. This may be especially helpful for some risky choice datasets where the stimuli are uniformly sampled from the probabilities rather than the odds against. In such case, fitting CBS in the odds against space may be suboptimal as the choices will be spread out unevenly over the odds space. In such circumstances, we recommend fitting CBS in the space of $1 - p$ and then transforming them into the odds space if needed. Furthermore, fitting CBS in the space of $1 - p$ can also easily show granular risk aversion for a given probability by checking if the fitted function is above the identity line $f(1 - p) = 1 - p$.

PAPER SUMMARY AND DISCUSSION

Cubic Bezier Splines are a promising method that can agnostically approximate utility functions, freeing researchers from justifying the selection and assumptions of a specific parametric utility model or from gathering large amounts of structured data for non-parametric estimation. CBS's history in computer graphics attests to its flexibility and ease-of-use. We have shown that CBS can closely approximate many of the established models, as well as unconventional behavioral patterns that other models have difficulty fitting. We also demonstrated that CBS outperforms all other models tested in cross-validation, therefore providing improved predictive accuracy over current approaches. Lastly, we showed that CBS picks up heterogeneity and granular patterns of choice that are consistent across time and cannot be captured by traditional models. In ITC, we showed that there are a group of participants that exhibit increasing discount rates that cannot be captured by the extant models we tested. In risky choice, we showed that each participant has a rich granular pattern of risk-aversion that is consistent across time. All these benefits provide compelling evidence that CBS can be a 'one-model-fits-all' approach to utility modeling that not only provides high flexibility and predictive power, but also yields significant new insights that are easy to interpret.

We also believe that CBS can further advance our understanding of human choice behavior. An important and understudied question involves variability in an individual's choices across time, and CBS' better estimates of preference functions will allow one to better estimate such choice variability without contamination from function misspecification. Better utility estimates will also aid current efforts to relate such utilities to response times, eye movements, or neural activity (Levy & Glimcher, 2012; Venkatraman, Payne, & Huettel, 2014). Finally, an important question for future work is whether the heterogenous choice patterns uncovered by CBS arise because different groups of participants use different sets of psychological processes

when making decisions (Reeck, Wall, & Johnson, 2017).

As a newly developed method, CBS has much potential for improvement. For one, it can be used under a Bayesian framework to obtain a posterior distribution of fitted utility functions, to allow hierarchical influence to aid model fitting, and to use priors to guide the model outside the range of sampled data. For example, one can impose priors on the empirically calculated moment-by-moment discount rates of the CBS to reduce overfitting and to help keep CBS in a reasonable shape outside the range of sampled data. Alternatively, one may also consider modern ensemble methods such as bagging and boosting to augment CBS's flexibility while fighting overfitting. One can also further use the possible constraints to CBS to better tailor it for specific usage. We have proven conditions for convexity and concavity in the web appendix but have not implemented these constraints in this paper since constraints beyond monotonicity were not empirically deemed necessary. Finally, future studies may develop adaptive questioning schemes that are based on CBS estimations to provide an efficient agnostic estimation of utility functions.

As we provide CBS as a new tool for describing, understanding, and predicting decisions, we hope that this research is the start of using a flexible, agnostic models to explore many topics not only related to economic decision-making tasks, but also in other cognitive, affective, and social decisions whose models have latent variables. We hope that across many areas of human decision making, the behavioral patterns and heterogeneity that went unnoticed under formal parametric assumptions can now easily be brought to surface and studied.

REFERENCES

Abdellaoui, M. (2000). Parameter-Free Elicitation of Utility and Probability Weighting Functions. *Management*.

Alessi, S. M., & Petry, N. M. (2003). Pathological gambling severity is associated with impulsivity in a delay discounting procedure. *Behavioural Processes*, *64*(3), 345–354.

Anderson, L. R., & Mellor, J. M. (2008). Predicting health behaviors with an experimental measure of risk preference. *Journal of Health Economics*, *27*(5), 1260–1274.

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

Brañas-Garza, P., Georgantzís, N., & Guillén, P. (2007). Direct and indirect effects of pathological gambling on risk attitudes. *Judgment and Decision Making*, *2*(2), 126–136.

Bruhin, A., Fehr-Duda, H., & Epper, T. (2009). Risk and Rationality: Uncovering Heterogeneity in Probability Distortion. *Ssrn*, *78*(4), 1375–1412.

De Casteljau, P. (1963). Courbes et surfaces à pôles. *André Citroën, Automobiles SA, Paris*.

Doyle, J. R. (2010). Survey of Time Preference, Delay Discounting Models. *Ssrn*.

Fox, C. R., & Poldrack, R. A. (2009). Prospect theory and the brain. In *Neuroeconomics* (pp. 145–173).

Goldstein, W., & Einhorn, H. (1987). Expression theory and the preference reversal phenomena. *Psychological Review*.

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

Green, L., Fry, A. F., & Myerson, J. (1994). Research Report Discounting of Delayed

Rewards:A Life-Span Comparison. *Pschological Science*, *5*(1), 33–36.

Green, L., Myerson, J., & Ostaszewski, P. (1999). Amount of Reward Has Opposite Effect on the Discounting of Delayed and Prrobabilistic Outcomes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. American Psychological Association.

Herrnstein, R. J. (1981). Self-control as response strength. *Quantification of Steady-State Operant Behavior*, 3–20.

Kable, J. W., Caulfield, M. K., Falcone, M., McConnell, M., Bernardo, L., Parthasarathi, T., … Lerman, C. (2017). No Effect of Commercial Cognitive Training on Brain Activity, Choice Behavior, or Cognitive Performance. *The Journal of Neuroscience*, *37*(31), 7390–7402.

Killeen, P. R. (2009). An Additive-Utility Model of Delay Discounting. *Psychological Review*.

Kirby, K. N., Petry, Nancy, M., & Bickel, Warren, K. (1999). Heroin addicts have higher discount rates for delayed rewards than non drug using controls. *Journal of Experimental Psychology*, *128*(1), 78–87.

Krain, A. L., Gotimer, K., Hefton, S., Ernst, M., Castellanos, F. X., Pine, D. S., & Milham, M. P. (2008). A Functional Magnetic Resonance Imaging Investigation of Uncertainty in Adolescents with Anxiety Disorders. *Biological Psychiatry*, *63*(6), 563–568.

Laibson, D. (1997). Golden Eggs and Hyperbolic Discounting. *The Quarterly Journal of Economics*, *112*(2), 443–478.

Lejuez, C. W., Aklin, W. M., Bornovalova, M. A., & Moolchan, E. T. (2005). Differences in risk-taking propensity across inner- city adolescent ever-and never-smokers. *Nicotine & Tobacco Reserach*, *7*(1), 71–79.

Lejuez, C. W., Aklin, W. M., Jones, H. A., Strong, D. R., Richards, J. B., Kahler, C. W., & Read, J. P. (2003). The Balloon Analogue Risk Task (BART) differentiates smokers and

nonsmokers. *Experimental and Clinical Psychopharmacology*, *11*(1), 26–33.

Levy, D. J., & Glimcher, P. W. (2012). The root of all value: A neural common currency for choice. *Current Opinion in Neurobiology*, *22*(6), 1027–1038.

Markowitz, H. (1959). Portfolio Selection, Cowles Foundation Monograph No. 16. *John Wiley, New York. S. Moss (1981). An Economic Theory of Business Strategy, Halstead Press, New York. TH Naylor (1966). The Theory of the Firm: A Comparison of Marginal Analysis and Linear Programming. Southern Economic Journal (January)*, *32*, 263–274.

Myerson, J., Green, L., & Warusawitharana, M. (2001). Area under the curve as a measure of discounting. *Journal of the Experimental Analysis of Behavior*.

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

Rachlin, H., Raineri, A., & Cross, D. (1991). Subjective probability and delay. *Journal of the Experimental Analysis of Behavior*, *55*(2), 233–244.

Reeck, C., Wall, D., & Johnson, E. J. (2017). Search predicts and changes patience in intertemporal choice. *Proceedings of the National Academy of Sciences*.

Samuelson, P. a. (1937). Note on Measurement of Utility. *The Review of Economic Studies*, *4*(2), 155–161.

Schepis, T. S., McFetridge, A., Chaplin, T. M., Sinha, R., & Krishnan-Sarin, S. (2011). A pilot examination of stress-related changes in impulsivity and risk taking as related to smoking status and cessation outcome in adolescents. *Nicotine and Tobacco Research*, *13*(7), 611–615.

Shamosh, N. A., & Gray, J. R. (2008). Delay discounting and intelligence: A meta-analysis. *Intelligence*, *36*(4), 289–305.

Stott, H. P. (2006). Cumulative prospect theory ' s functional menagerie. *Journal of Risk and*

Uncertainty, *32*(2), 101–130.

Tjur, T. (2009). Coefficients of determination in logistic regression models - A new proposal: The coefficient of discrimination. *American Statistician*.

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

Venkatraman, V., Payne, J. W., & Huettel, S. A. (2014). An overall probability of winning heuristic for complex risky decisions: Choice and eye fixation evidence. *Organizational Behavior and Human Decision Processes*, *125*(2), 73–87.

Wakker, P., & Deneffe, D. (1996). Eliciting von Neumann-Morgenstern Utilities When Probabilities Are Distorted or Unknown. *Management Science*, *42*(8), 1131–1150.

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.

FOOTNOTES

1. One possible alternative to assuming one specific parametric model is to employ a mixture model so that participants are fitted by the best fitting model. However, mixture models also assume that one of the entertained specific models is true.

2. There is the original prospect theory model and the cumulative prospect theory model whose forms and interpretations differ. In this manuscript, we only deal with gambles with two or less outcomes, in which case their mathematical form is identical.

3. Technically, these agnostic models can be parametric or non-parametric. However, to disambiguate from formal parametric models that assume a rigid structure (e.g., eq. 1,2,3,4), we refer to these methods as agnostic.

4. This definition makes CBS a parametric curve, but not necessarily a parametric function of $y = f(x)$. CBS is parametric in the sense that a cardioid or a circle is also parametric.

5. The models shown here typically assumes a linear value function $f(A) = A$. If this is not met, CBS's modeling of the discounting function would combine the effect of both $f(A)$ and $g(D)$. See risky choice section for such examples.

6. We are simulating in original form but comparing the CBS fit against discounting form to clearly show that the prospect theory forms are analytically convertible to discounting forms.

7. Although it makes no difference, the models were fitted in their original forms (as opposed to the discounting form).

TABLES

ITC	Exponential		Hyperbolic		Gen. Hyp.		Quasi-hyp		Nonparam	
# Choice	CBS1	CBS2	CBS1	CBS2	CBS1	CBS2	CBS1	CBS2	CBS1	CBS2
49	**.031**	.035	**.036**	.040	**.036**	.040	.030	**.030**	.072	**.040**
64	**.029**	.031	**.031**	.035	**.032**	.036	.030	**.030**	.068	**.039**
81	**.027**	.027	**.027**	.031	**.029**	.032	.028	**.024**	.068	**.032**
100	**.025**	.026	**.025**	.028	**.026**	.030	.024	**.022**	.063	**.031**
121	**.024**	.024	**.021**	.026	**.023**	.027	.023	**.021**	.054	**.028**
144	.022	**.022**	**.021**	.025	**.022**	.025	.022	**.016**	.062	**.027**
169	.021	**.019**	**.019**	.023	**.020**	.024	.021	**.016**	.056	**.023**
196	.020	**.017**	**.018**	.022	**.019**	.023	.020	**.015**	.061	**.025**
225	.020	**.017**	**.017**	.021	**.017**	.021	.020	**.014**	.057	**.021**
256	.019	**.016**	**.016**	.019	**.017**	.020	.019	**.013**	.052	**.020**
289	.018	**.015**	**.015**	.019	**.015**	.019	.018	**.012**	.057	**.020**
324	.017	**.015**	**.014**	.018	**.015**	.018	.018	**.012**	.049	**.020**
361	.017	**.014**	**.013**	.017	**.014**	.017	.017	**.011**	.056	**.018**
400	.017	**.013**	**.013**	.016	**.013**	.016	.017	**.010**	.048	**.018**

Table 1. Mean absolute errors (MAE) of CBS fits at various ITC dataset sizes. *For each simulating function group (i.e., exponential, hyperbolic, generalized hyperbolic, quasi-hyperbolic, and non-parametric), we show the MAEs of the 1-piece and 2-piece CBS. The lower MAEs of the two CBS fits are shown bolded.*

Risky choice	EUT		Tversky		Goldstein		Prelec	
# Choice	CBS1	CBS2	CBS1	CBS2	CBS1	CBS2	CBS1	CBS2
49	**.050**	.059	**.044**	.051	**.040**	.053	**.038**	.052
64	**.040**	.048	**.036**	.040	**.033**	.042	**.032**	.041
81	**.034**	.043	**.031**	.034	**.029**	.037	**.027**	.035
100	**.030**	.038	**.027**	.029	**.025**	.032	**.024**	.032
121	**.027**	.033	**.026**	.027	**.021**	.028	**.023**	.030
144	**.025**	.032	**.026**	.026	**.020**	.027	**.021**	.027
169	**.023**	.029	.024	**.022**	**.018**	.023	**.018**	.025
196	**.022**	.027	.021	**.020**	**.017**	.023	**.017**	.023
225	**.019**	.025	.020	**.019**	**.016**	.021	**.016**	.021
256	**.018**	.024	.020	**.017**	**.015**	.020	**.015**	.020
289	**.017**	.023	.020	**.017**	**.014**	.018	**.014**	.019
324	**.016**	.021	.019	**.017**	**.013**	.018	**.014**	.018
361	**.016**	.020	.019	**.014**	**.013**	.017	**.013**	.017
400	**.015**	.019	.017	**.013**	**.012**	.016	**.012**	.016

Table 2. Mean absolute errors (MAE) of CBS fits at various risky choice dataset sizes. *For each simulating group (i.e., EUT, Tversky-Kahneman, Goldstein-Einhorn, Prelec), we show the MAEs of the 1-piece and 2-piece CBS. The lower MAEs of the two CBS fits are shown bolded.*

FIGURES

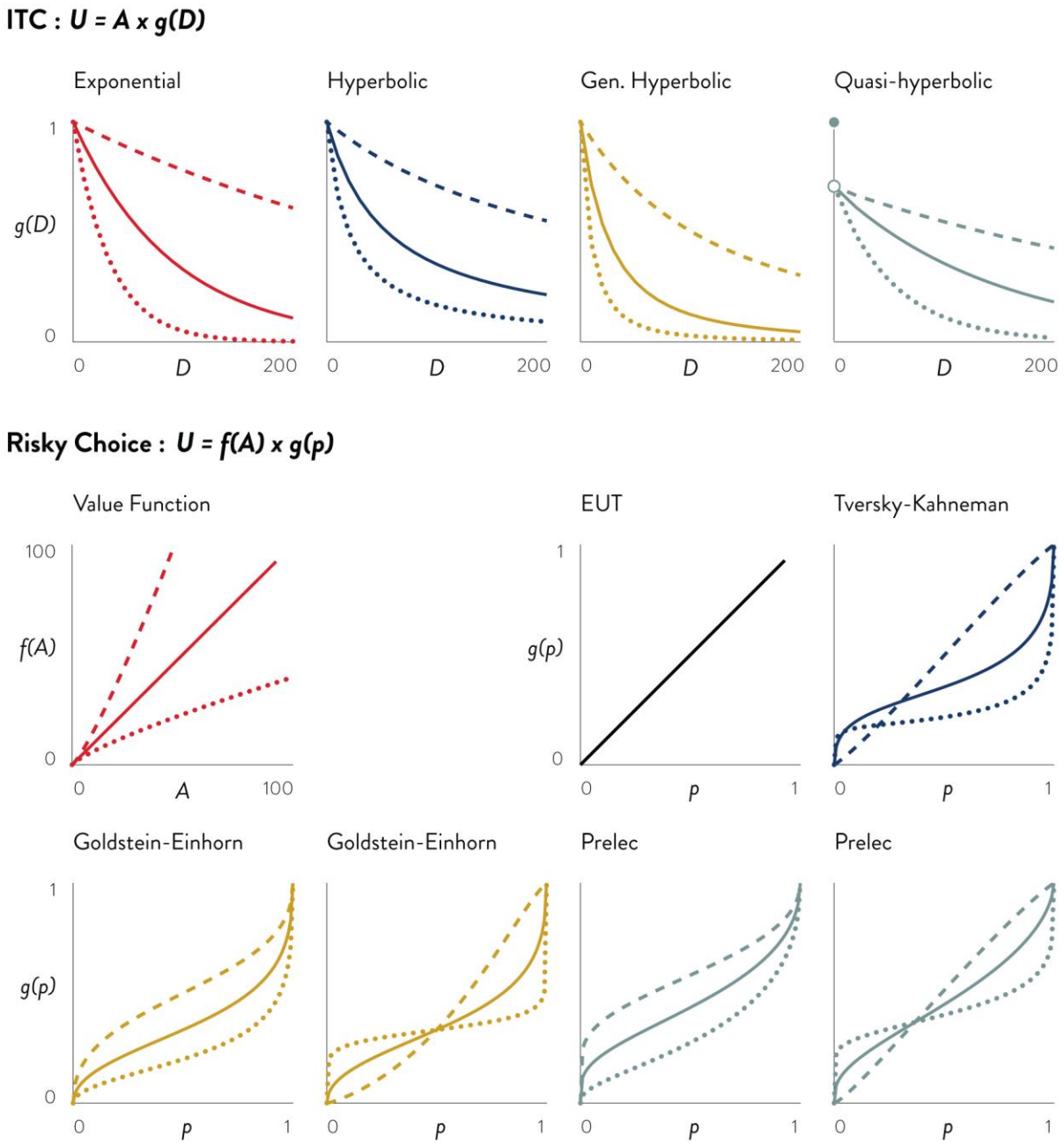


Figure 1. Various utility functions for ITC and risky choice represented by its latent variable transformations. *Various parameter combinations have been selected to show the range of shapes. Exponential (eq. 1) : ln(k) = -6, -4.5, -3.5. Hyperbolic (eq. 3a) : ln(k) -5.5, -4, -3. Generalized hyperbolic (eq. 3c) : same as hyperbolic with s = 2 for all. Quasi-hyperbolic (eq. 3b) : ln(k) = -6, -5, -4 with β = 0.7 for all. Risky choice f(A) = A^{α}, with α = 1.2, 1, 0.8. Goldstein-Einhorn (eq. 5a) : (γ, δ) = (0.5, 0.25), (0.5, 0.5), (0.5, 1), (0.15, 0.5), (0.5, 0.5), (1.2, 0.5). Tversky-Kahneman (eq. 4c) : γ = 0.35, 0.5, 1.1. Prelec (eq. 5b) : (γ, δ) = (0.5, 0.65), (0.5, 1), (0.5, 1.55), (0.25, 1), (0.6, 1), (1.1, 1).*

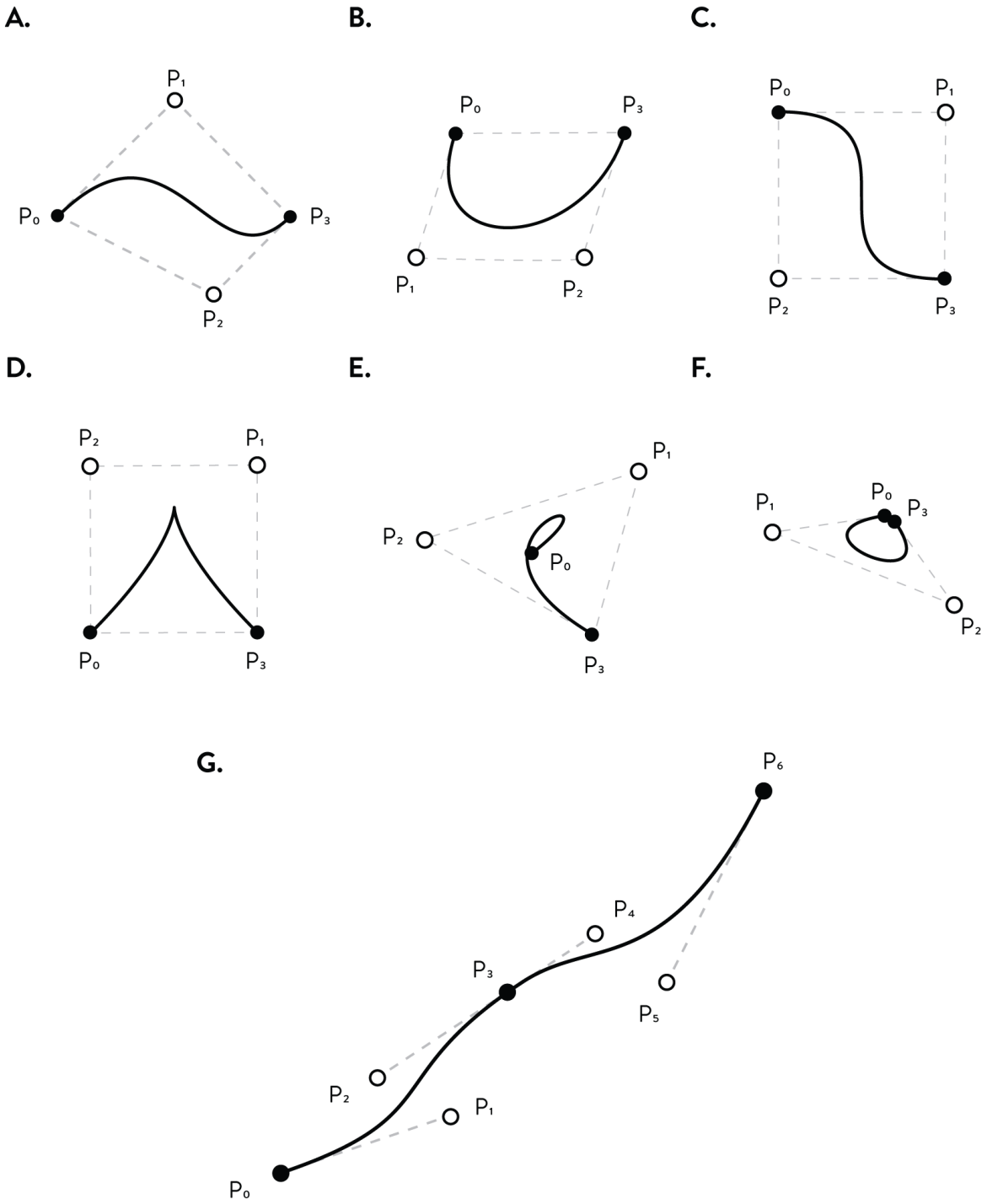


Figure 2. Various examples of cubic Bezier splines (black line) and their control points (black dots). *A cubic Bezier spline starts from the first control point and ends at the last control point while being pulled towards the other two control points. Due to its convex hull property, the entire spline is bounded by the outermost polygon drawn by the control points. A through F shows various 1-piece CBS splines bounded by their convex hill. G shows 2-piece CBS spline smoothly joined together. The first piece is on the left side, drawn by the four points P_0~P_3, and the second piece is on the right side, drawn by the four points P_3~P_6.*

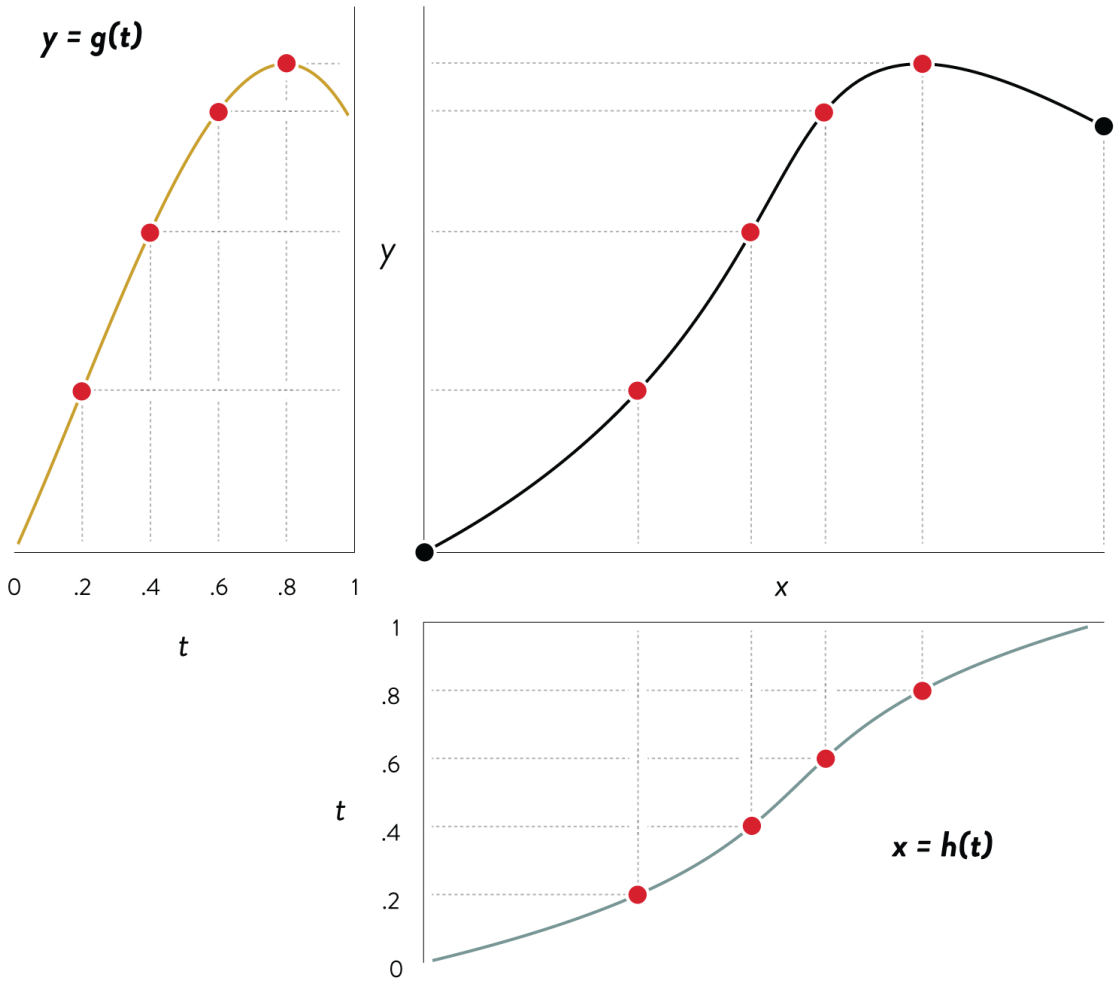


Figure 3. Relationship between CBS and its parametric functions x = h(t) and y = g(t). *An example CBS is drawn on the right upper panel in black. The two functions x = h(t) and y = g(t) is drawn on the left and the bottom to show the correspondence between t and x and y. The dots indicate the coordinates of CBS at t = 0.2, 0.4, 0.6, and 0.8.*

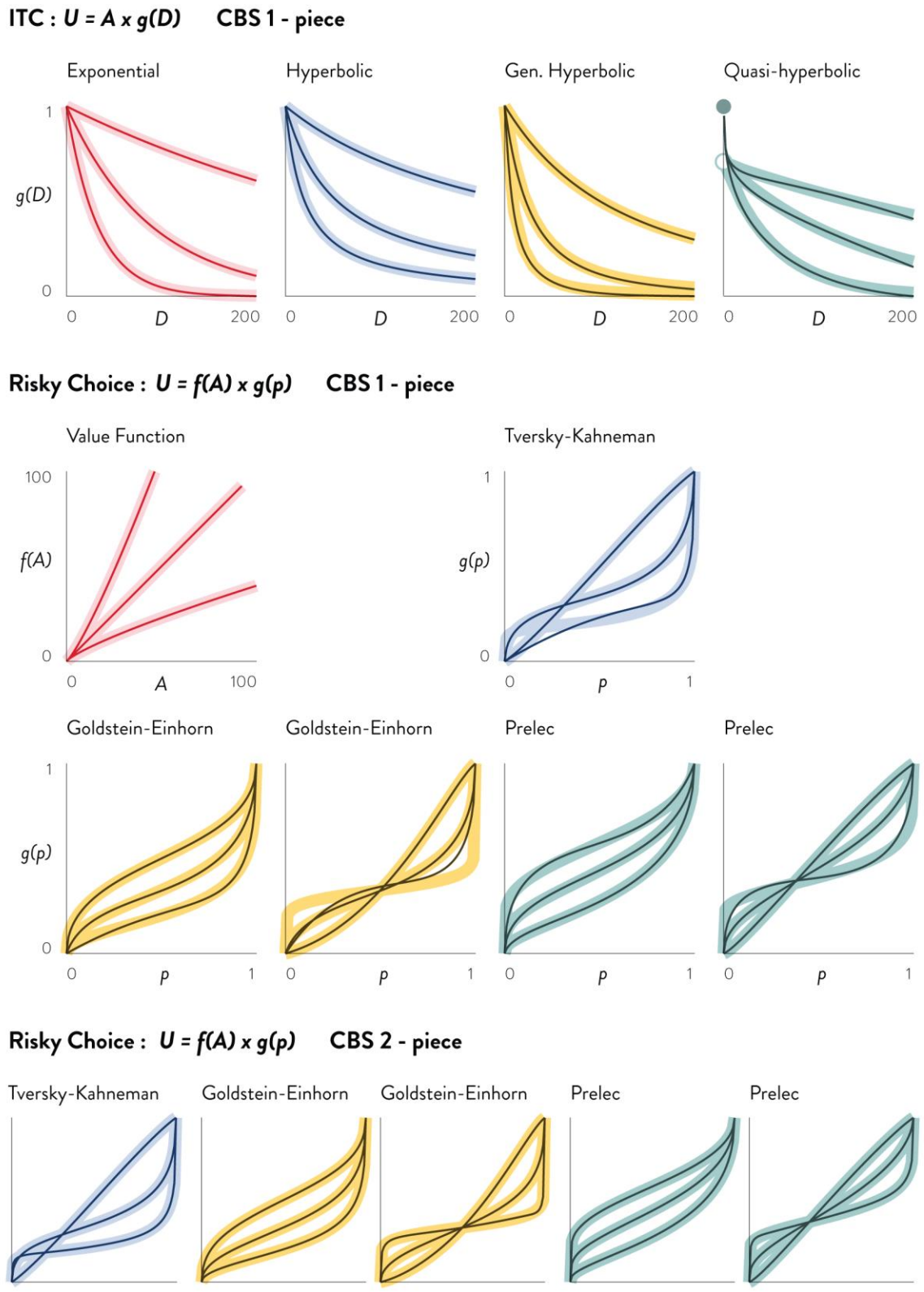


Figure 4. Examples of CBS fits to parametric utility functions. *The parameters of each utility function is the same as in Figure 1. The broad lighter lines are the true functions while the thin darker lines are the fitted CBS functions.*

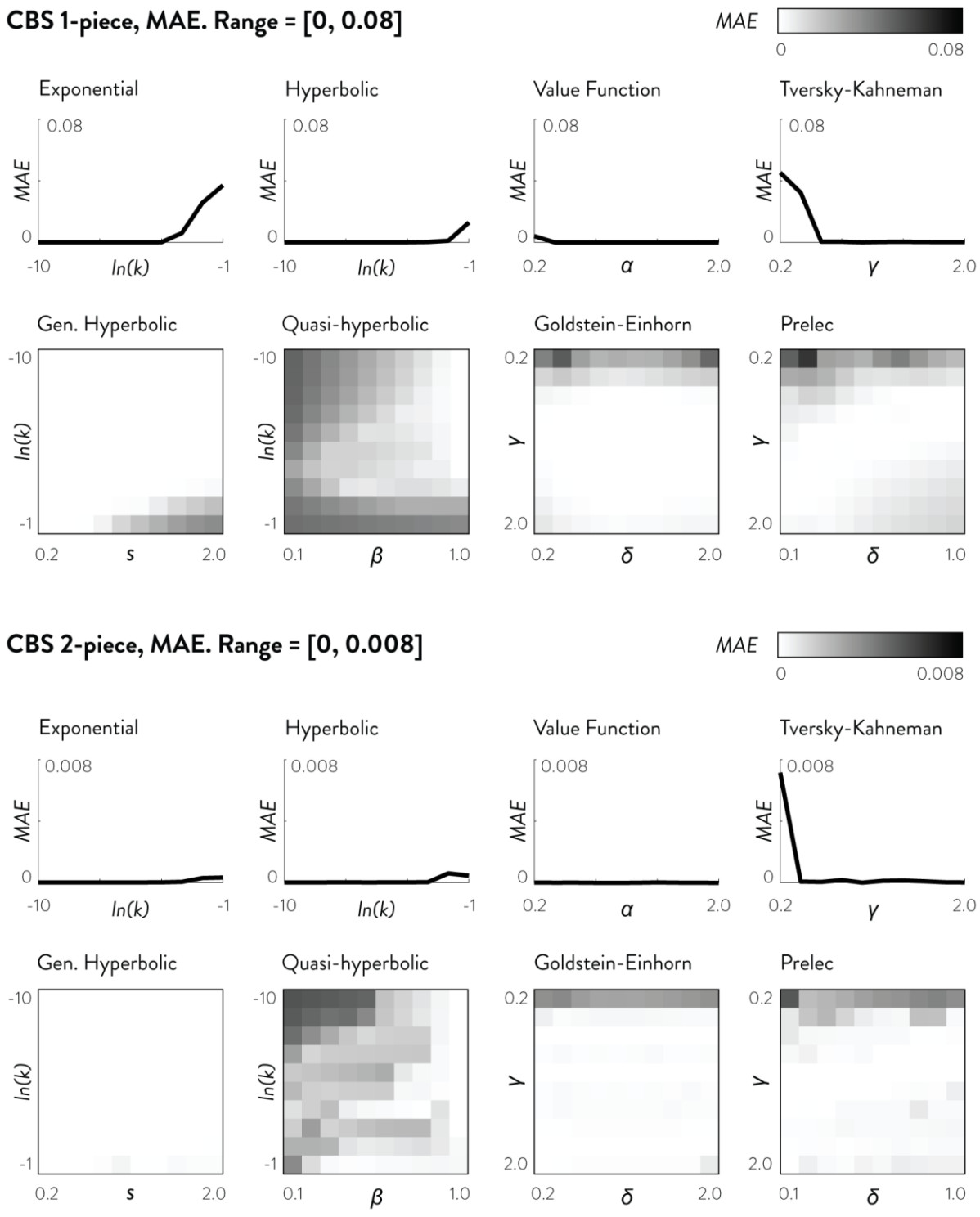


Figure 5. Analysis of mean absolute error of CBS fits to parametric utility functions. *The top half shows MAE of 1-piece CBS fits and the bottom half shows MAE of 2-piece CBS fits. MAE of 1-parameter functions are shown in line plots (ITC: exponential, hyperbolic; risky choice: value function, Tversky-Kahneman). MAE of 2-parameter functions are shown in heatmaps (ITC: generalized hyperbolic, quasi-hyperbolic; risky choice: Goldstein-Einhorn, Prelec).*

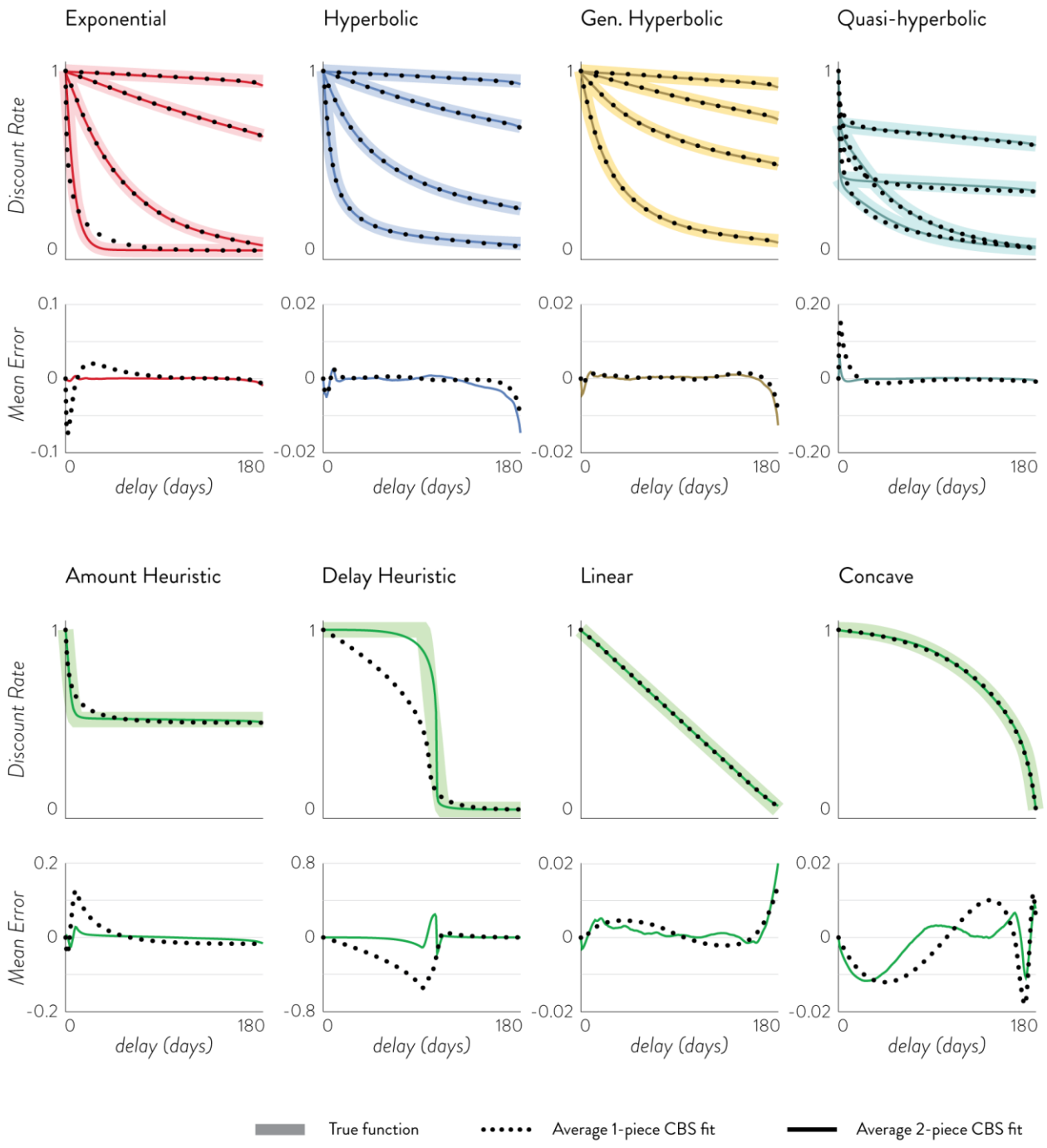


Figure 6. Average CBS fits from ITC choice dataset simulation. *The top two rows show the 16 parametric simulations and the bottom two rows show the 4 non-parametric simulations. The first and the third row shows the true simulating utility functions and their average CBS fits. The average CBS fits were calculated by taking the mean of the 1000 fitted CBS functions from the largest choice dataset (400 choices). The second and the fourth row shows the mean error of the fitted CBS functions and the true simulating functions.*

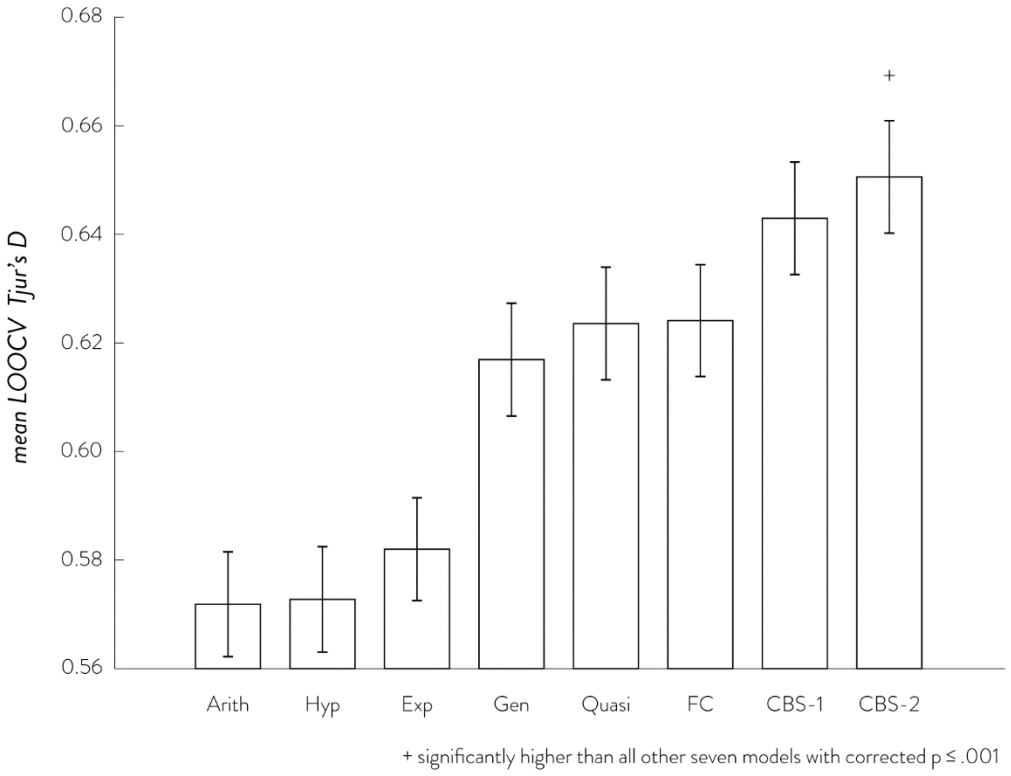


Figure 7. Mean leave-one-out cross-validation Tjur's D of 8 models of ITC. *The error bars are standard errors of the mean. From left to right, the models are arithmetic, hyperbolic, exponential, generalized hyperbolic, quasi-hyperbolic, fixed cost, 1-piece CBS, and 2-piece CBS.*

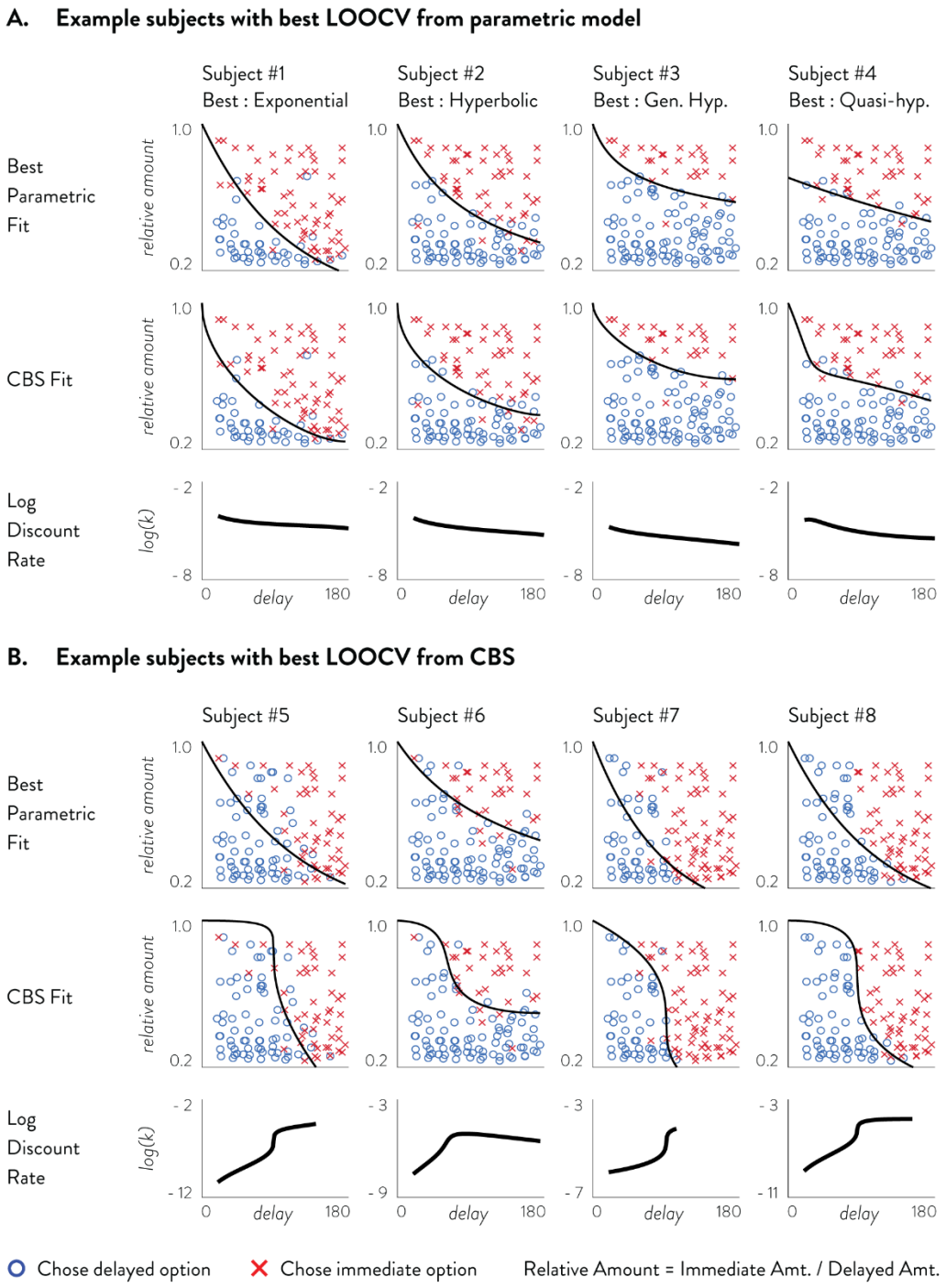


Figure 8. Plots of eight example participants' choices, their best parametric fits, their 1-piece CBS fits, and their moment-by-moment log discount rate calculated from 1-piece CBS fits. *Panel A shows 4 participants whose highest LOOCV accuracy came from parametric models and panel B shows 4 participants whose highest LOOCV accuracy came from CBS. In each panel, the top row shows the best parametric model (by LOOCV) and the middle row shows the 1-piece CBS fit. The bottom row shows the moment-by-moment log discount rates. Top panel participants were selected such that the full curve would be visually shown; bottom panel participants were selected to show variety of CBS fits that did not conform to parametric forms.*

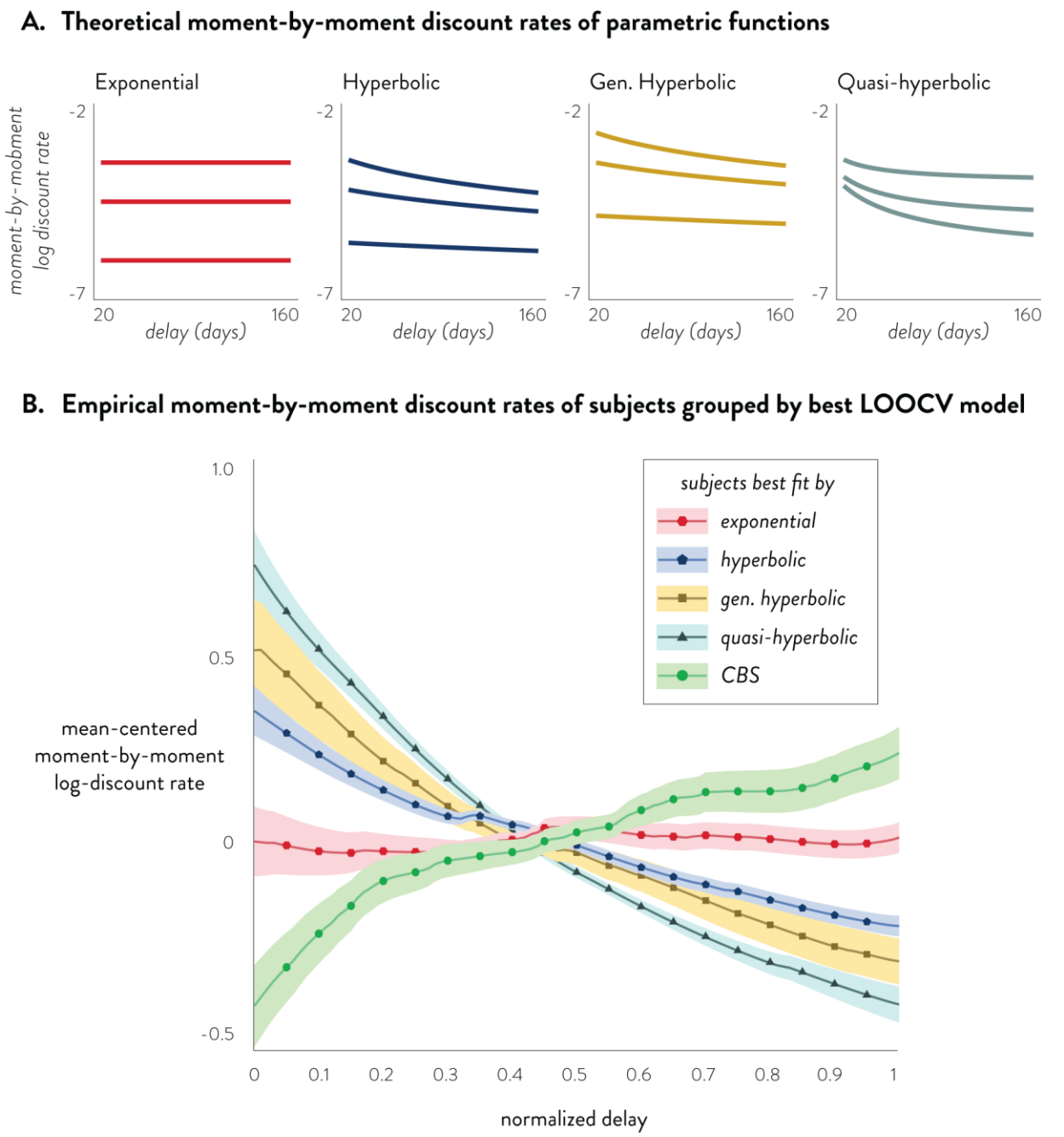


Figure 9. Theoretical moment-by-moment discount rates (A) and empirical average mean-centered moment-by-moment log discount rates of all 285 session's data grouped by models of highest LOOCV accuracy (B). *Panel A shows theoretical multiplicative discount rates of 4 parametric models shown in figure 1 and 5. Exponential models have constant discount rates over time whereas other parametric models have decreasing discount rates. Panel B shows the empirical moment-by-moment discount rates calculated from CBS fits to real data. 74 Sessions had highest LOOCV accuracy with exponential, 49 with hyperbolic, 50 with generalized hyperbolic, 14 with quasi-hyperbolic, and 98 with CBS (50 with 1-piece and 48 with 2-piece).*

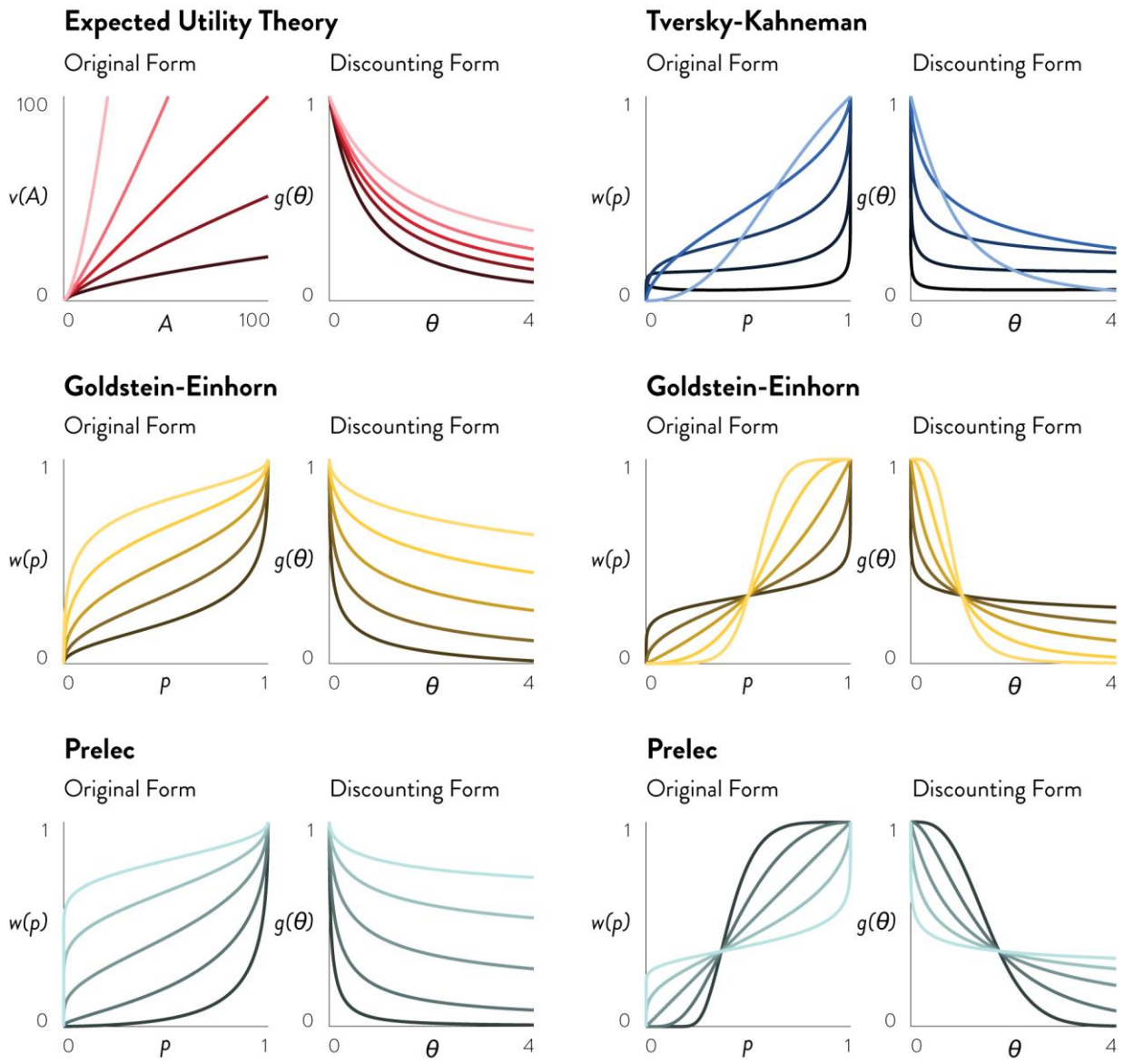


Figure 10. Analytical conversion of EUT and prospect theory models into discounting form. *Each of the 6 panels show the original form of function on the left and their color-coded converted discounting form on the right. For EUT, we used* $\alpha = 0.66, 0.85, 1, 1.17, 1.5$. *All other functions held* α *constant at 1. For Tversky-Kahneman, we used* $\gamma = 0.2, 0.3, 0.45, 0.7, 2$. *For Goldstein-Einhorn on the left, we fixed* $\gamma = 0.5$ *and varied* $\delta = 0.25, 0.5, 1, 2, 4$. *For Goldstein-Einhorn on the right, we fixed* $\delta = 0.5$ *and varied* $\gamma = 0.2, 0.5, 1, 2, 4$. *For Prelec on the left, we fixed* $\gamma = 0.5$ *and varied* $\delta = 0.25, 0.5, 1, 2, 4$. *For Prelec on the right, we fixed* $\delta = 1$ *and varied* $\gamma = 0.2, 0.5, 1, 2, 4$. *Since the right Prelec function has* δ *fixed at 1, it is the 1-parameter version of the Prelec weighting function.*

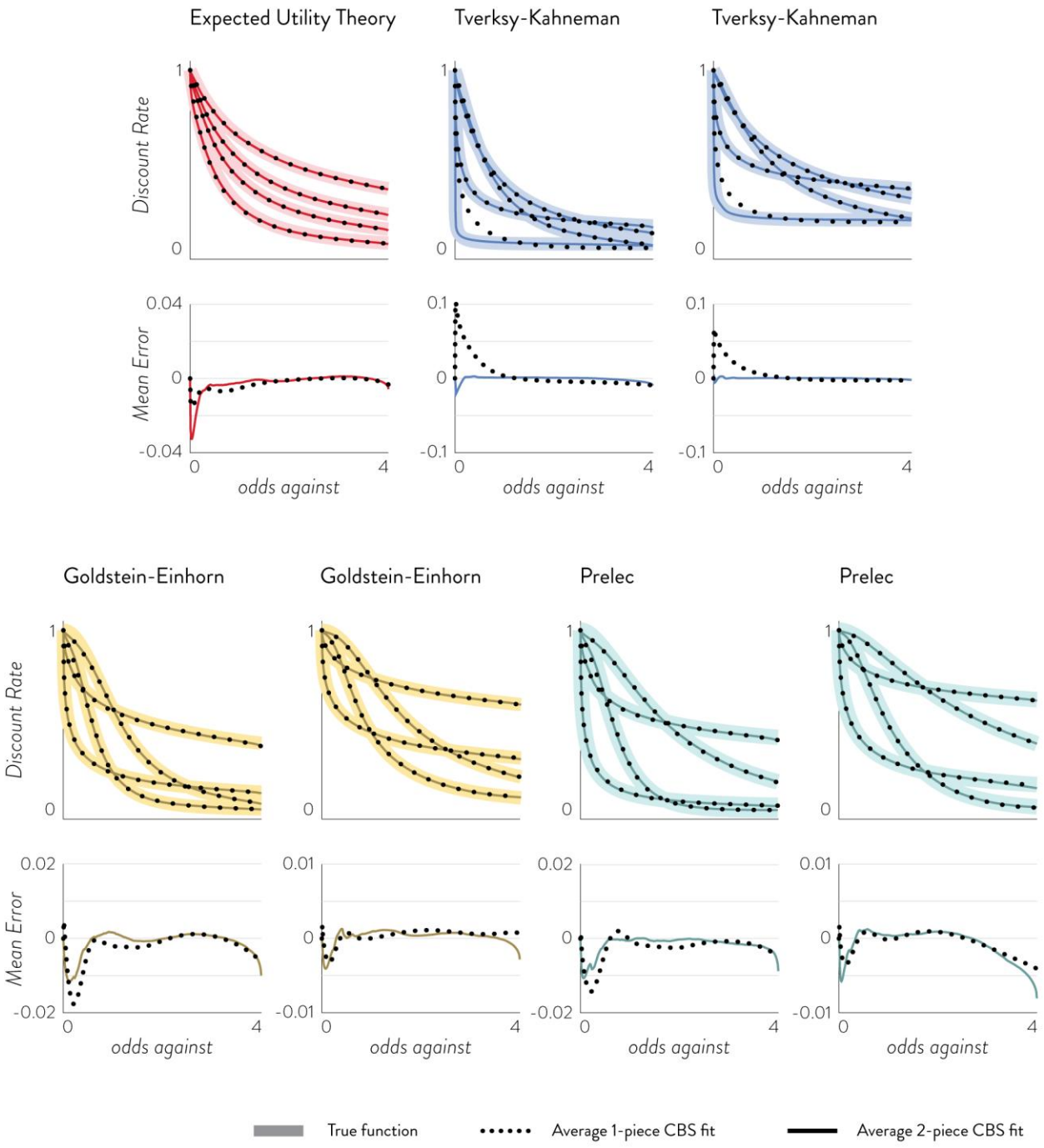


Figure 11. Average CBS fits from risky choice dataset simulation. *CBS fits are shown overlaid on top of the discounting form of prospect theory models. The first and the third rows show the true simulating utility functions and their average CBS fits. The average CBS fits were calculated by taking the mean of the 1000 fitted CBS functions from the largest choice dataset (400 choices). The second and the fourth rows show the mean error of the fitted CBS functions and the true simulating functions.*

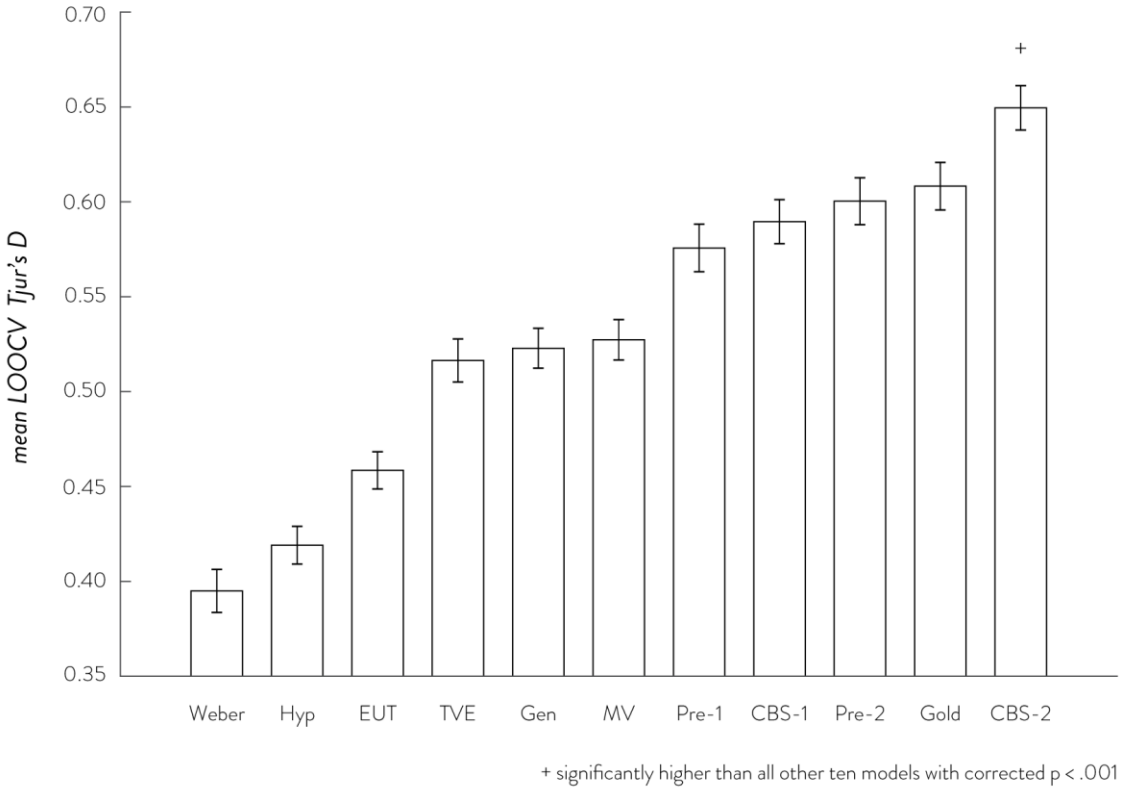


Figure 12. Mean leave-one-out cross-validation Tjur's D of 11 models of risky choice. *The error bars are standard errors of the mean. From left to right, the models are Weber's coefficient of variation, hyperbolic probability discounting, expected utility theory, Tverksy-Kahneman prospect theory, generalized hyperbolic probability discounting, mean variance, 1-parameter Prelec prospect theory, 1-piece CBS, 2-parameter Prelec prospect theory, Goldstein-Einhorn prospect theory, and 2-piece CBS.*

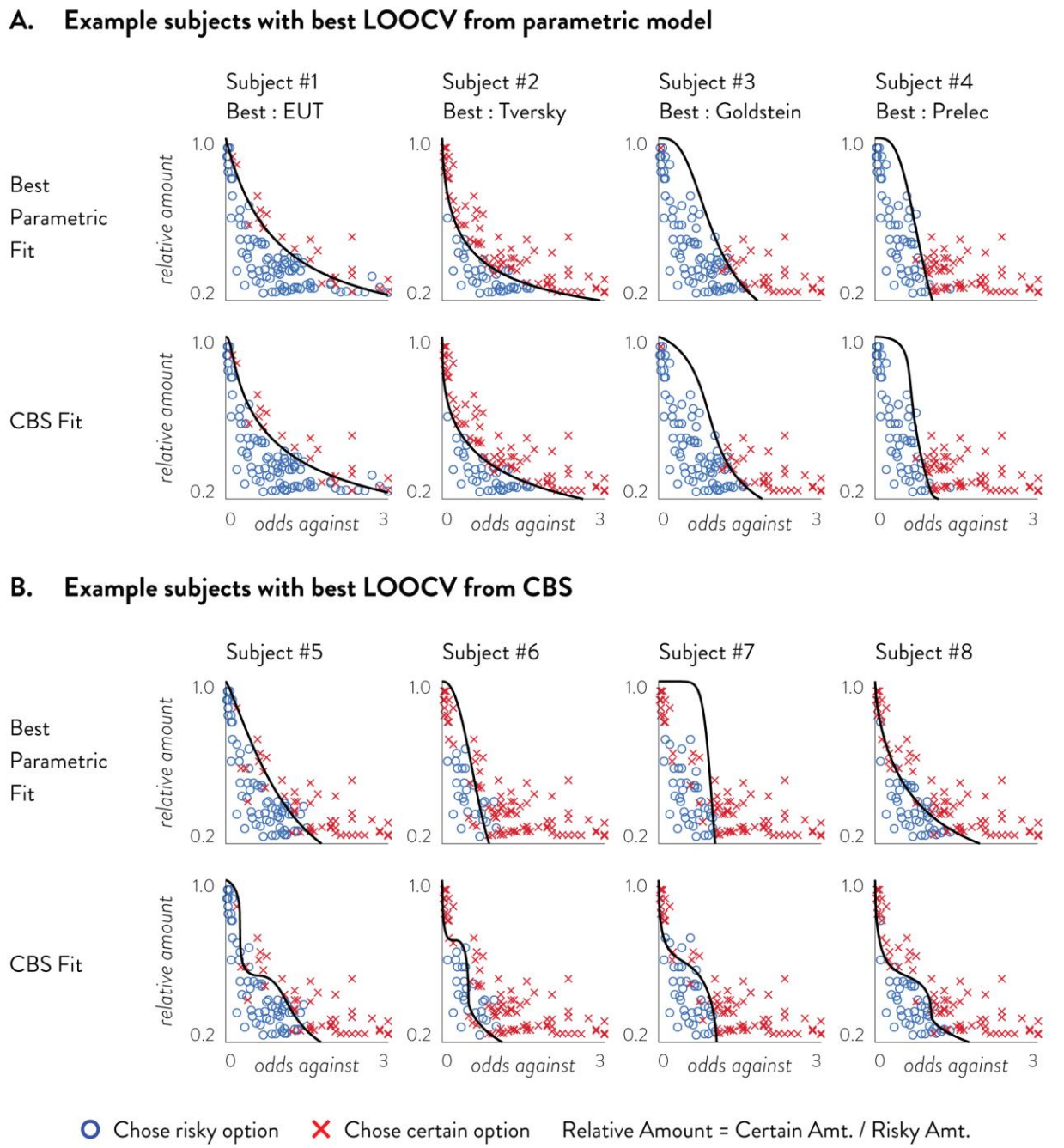


Figure 13. Plots of eight example participants' choices, their best parametric fits, and their 2-piece CBS fits. *Panel A shows 4 participants whose highest LOOCV accuracy came from parametric models and panel B shows 4 participants whose highest LOOCV accuracy came from CBS. In each panel, the top row shows the best parametric model (by LOOCV) and the bottom row shows the 2-piece CBS fit. Top panel participants were selected such that the distinctive full curves would be visually shown; bottom panel participants were selected to show variety of CBS fits that did not conform to extant parametric forms.*

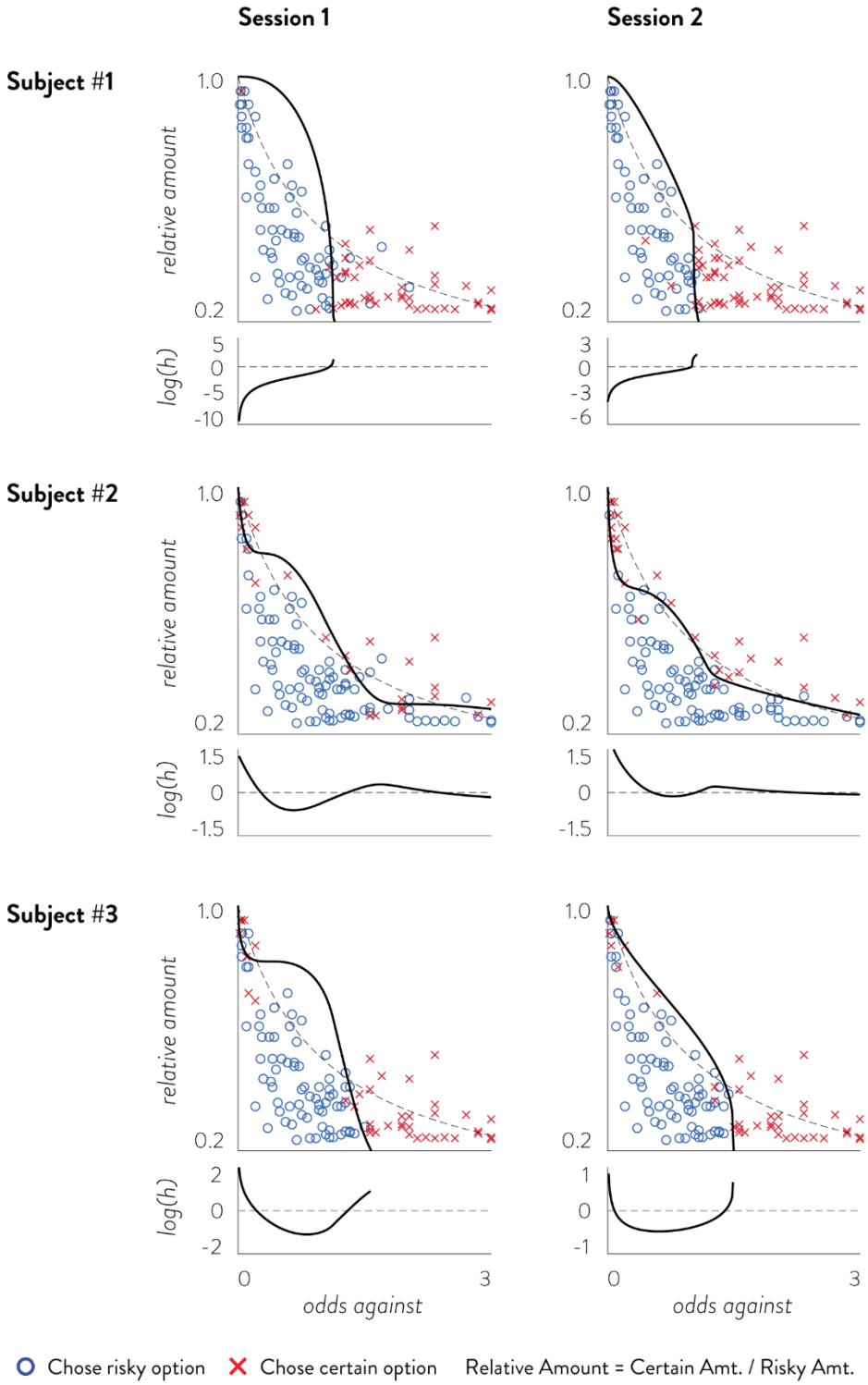
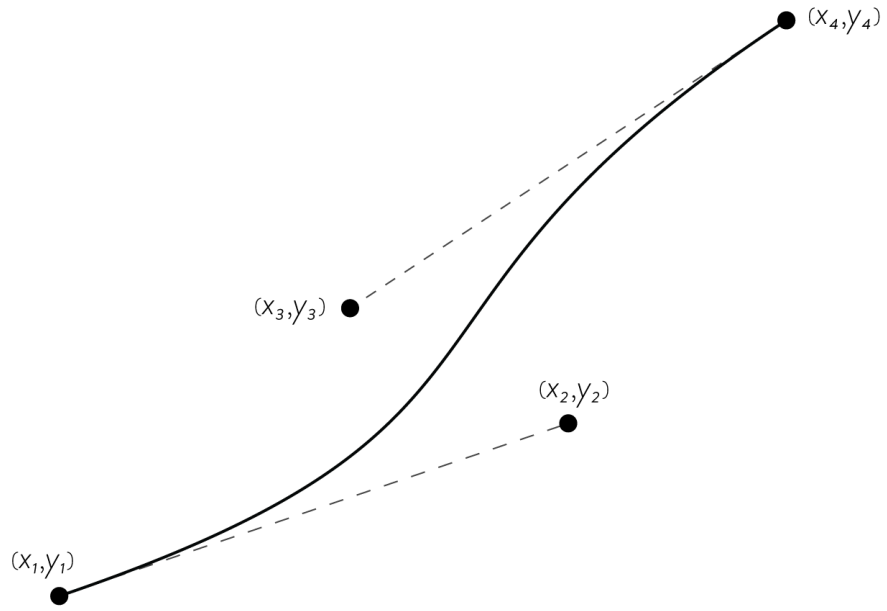


Figure 14. Example participants' cross-session similarity in risky choices and granular measure of risk aversion. *Three example participants' data are shown (one in each row) for both sessions (left and right columns). Their choices and its 2-piece CBS fits are shown on the top with a granular moment-by-moment measure of risk aversion is shown below. These participants were selected to show diverse patterns of moment-by-moment risk aversion patterns*

WEB APPENDIX

1. Compulsory and Slope Conditions of CBS



Supplement Figure 4. A single piece of CBS. *A single piece of CBS function is created by four points, two of which are called anchor points and the other two are called control points. The CBS is essentially a polynomial interpolation between these four points such that CBS starts from one anchor point and ends at another anchor point while being pulled towards the two control points. For this reason, CBS always passes through the anchor points (point 1 and 4 above) while it usually doesn't pass through the control points (point 2 and 3 above).*

Because chains of CBS are locally adjustable, the derivation of constraints only need to be worked out with regards to one CBS. A single piece of CBS is described by the following two parametric curves:

$$x(t) = (1-t)^3 x_1 + 3(1-t)^2 t x_2 + 3(1-t)t^2 x_3 + t^3 x_4, \qquad 0 \le t \le 1 \tag{1}$$

$$y(t) = (1-t)^3 y_1 + 3(1-t)^2 t y_2 + 3(1-t)t^2 y_3 + t^3 y_4, \qquad 0 \le t \le 1 \tag{2}$$

In order to use CBS to approximate a function of the form $y = f(x)$, we can assume $x_1 < x_4$ in order to have a spline of non-zero length. Also, we must ensure that $x(t)$ is a monotonically increasing function of t in $[0 \;\; 1]$. Otherwise, multiple values of y may exist for one x.

One may also want to impose additional constraints on CBS. In terms of the first derivative, one can make it monotonically increasing or monotonically decreasing. Given that $x(t)$ is a monotonically increasing function of t, it is only necessary to control for $y(t)$ for this slope constraint. One may also want to constrain CBS with the sign of the second derivative to be concave or convex.

The derivative of $x(t)$ with regards to t is as follows:

$$\frac{dx}{dt} = 3\{(-x_1 + 3x_2 - 3x_3 + x_4)t^2 + 2(x_1 - 2x_2 + x_3)t - x_1 + x_2\} \tag{3}$$

To ensure that dx/dt is positive in $[0\ \ 1]$: 1) dx/dt is positive at $t = 0$ and $t = 1$, and 2) $dx/dt = 0$ has no real roots in $[0\ \ 1]$. The first condition gives the following two inequalities:

$$x_1 < x_2, \qquad x_3 < x_4 \tag{4}$$

For the second condition, we employ a monotonic transformation of $z = t/(1-t)$, in which case $0 < t < 1$ translates to $0 < z$. Then we only need to ensure that $dx/dz = 0$ does not have any positive real roots. After conversion and arrangement, $dx/dz = 0$ becomes the following:

$$(x_3 - x_4)z^2 + 2(x_2 - x_3)z + x_1 - x_2 = 0 \tag{5}$$

Since $x_3 \neq x_4$, equation (5) is quadratic and the roots are the following:

$$z = \frac{(x_3 - x_2) \pm \sqrt{(x_3 - x_2)^2 - (x_2 - x_1)(x_4 - x_3)}}{x_3 - x_4} \tag{6}$$

It must be that either the determinant is negative or that it is non-negative but the roots are negative. In order for the determinant to be negative, the following must be true:

$$-\sqrt{(x_4 - x_3)(x_2 - x_1)} < x_3 - x_2 < \sqrt{(x_4 - x_3)(x_2 - x_1)} \tag{7}$$

If the determinant is non-negative, the converse is true:

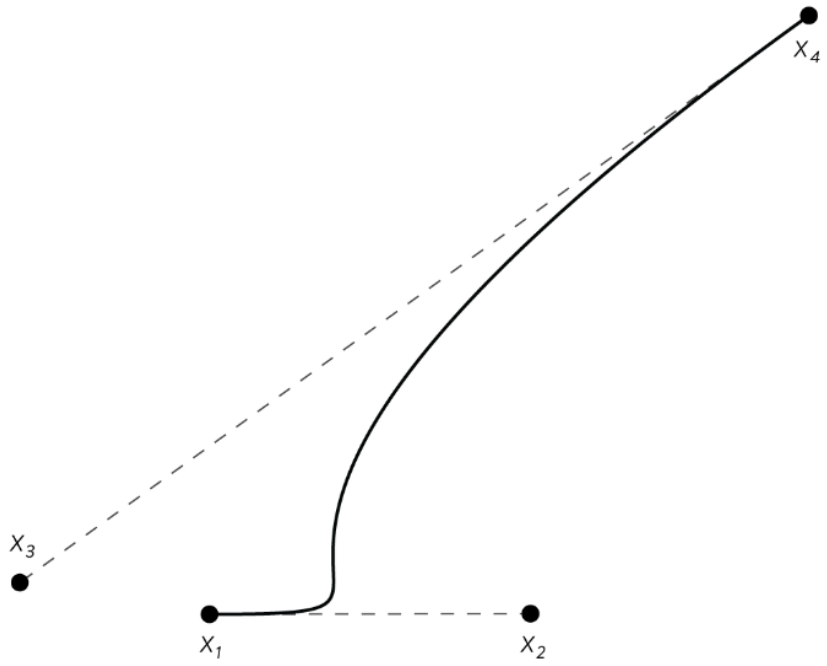
$$x_3 - x_2 \leq -\sqrt{(x_4 - x_3)(x_2 - x_1)} \qquad or \qquad \sqrt{(x_4 - x_3)(x_2 - x_1)} \leq x_3 - x_2 \tag{8}$$

If the left part of inequality (8) is true, it means that $x_3 - x_2$ is negative, which leads to at least one root of z (equation 6) being positive (since the denominator is negative). Hence, only the right part of inequality (8) can be true.

Combining all our results so far, we have the following compulsory conditions:

$$x_1 < x_4, \qquad x_1 < x_2, \qquad x_3 < x_4, \qquad -\sqrt{(x_4 - x_3)(x_2 - x_1)} < x_3 - x_2 \tag{9}$$

While these conditions guarantee that $x(t)$ is a monotonic function of t, we found that it is not ideal in two regards: 1) the final constraint is a non-linear inequality that entangles the 4 coordinates simultaneously, thereby reducing their independence, and 2) it allows a single piece of CBS to be "too flexible".



Supplement Figure 5. Kink created by a single piece of CBS.

As shown in Figure 2, we can see that even a monotonically increasing CBS can create a sharp bend, which are not ideal for preserving the smoothness of the function.

For this reason, we provide a more restrictive constraint of monotonicity that also allows the parameters to be more independent from each other:

$$x_1 < x_4, \qquad x_1 < x_2, \qquad x_3 < x_4, \qquad x_1 < x_3, \qquad x_2 < x_4 \tag{10}$$

These constraints make it so that as long as x_2 and x_3 stay within $[x_1, x_4]$, monotonicity is conserved. We can show that this satisfies the conditions on (9) by the following proof.

If $x_2 < x_3$, then $0 < x_3 - x_2$, which obviously satisfies $-\sqrt{(x_4 - x_3)(x_2 - x_1)} < x_3 - x_2$. If otherwise ($x_2 \geq x_3$), then it means that $x_1 < x_3 \leq x_2 < x_4$ under conditions in (10). Then, we can see that $(x_4 - x_3)(x_2 - x_1) = \{(x_4 - x_2) + (x_2 - x_3)\}\{(x_2 - x_3) + (x_3 - x_1)\} = (x_2 - x_3)^2 + (x_4 - x_2)(x_2 - x_3) + (x_3 - x_1)(x_2 - x_3) + (x_4 - x_2)(x_3 - x_1)$ and that this is strictly greater $(x_3 - x_2)^2$ since all terms are positive. Therefore, $-\sqrt{(x_4 - x_3)(x_2 - x_1)} < x_3 - x_2$ is satisfied.

Finally, we can see that the very first inequality in (10) is now obsolete because the other inequalities (e.g., $x_1 < x_2$, $x_2 < x_4$) already imply it. The constraint for slope is very similar to the compulsory condition as one only needs to swap x and y.

2. Curvature Conditions of CBS

Constraint for curvature must be done using the second derivative d^2y/dx^2:

$$\frac{d^2y}{dx^2} = \frac{\frac{d}{dt}\left(\frac{dy/dt}{dx/dt}\right)}{\frac{dx}{dt}} = \frac{\frac{d^2y}{dt^2}\frac{dx}{dt} - \frac{dy}{dt}\frac{d^2x}{dt^2}}{\left(\frac{dx}{dt}\right)^3} \tag{11}$$

Since our interest lies in constraining the sign of the 2nd derivative, the denominator is unnecessary for our purpose as it is always positive. The numerator becomes the following quadratic function of t:

$$pt^2 + qt + r \tag{12}$$

$p = (y_1 - 2y_2 + y_3)(x_1 - 3x_2 + 3x_3 - x_4) - (x_1 - 2x_2 + x_3)(y_1 - 3y_2 + 3y_3 - y_4)$

$q = (x_1 - x_2)(y_1 - 3y_2 + 3y_3 - y_4) - (y_1 - y_2)(x_1 - 3x_2 + 3x_3 - x_4)$

$r = (y_1 - y_2)(x_1 - 2x_2 + x_3) - (x_1 - x_2)(y_1 - 2y_2 + y_3)$

Let V_0 and V_1 denote the evaluation of equation 11 at $t = 0$ and $t = 1$:

$$\left.\frac{d^2y}{dx^2}\right|_{t=0} \propto r = V_0 = -x_2y_1 + x_3y_1 + x_1y_2 - x_3y_2 - x_1y_3 + x_2y_3 \tag{13}$$

$$\left.\frac{d^2y}{dx^2}\right|_{t=1} \propto p + q + r = V_1 = -x_3y_2 + x_4y_2 + x_2y_3 - x_4y_3 - x_2y_4 + x_3y_4 \tag{14}$$

The interpretation of V_0 and V_1 becomes clear when one considers their various forms:

$$\begin{aligned} V_0 &= -x_2y_1 + x_3y_1 + x_1y_2 - x_3y_2 - x_1y_3 + x_2y_3 \\ &= (x_2 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_1) \\ &= (x_2 - x_3)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_3) \\ &= (x_2 - x_3)(y_2 - y_1) - (x_2 - x_1)(y_2 - y_3) \end{aligned} \tag{15}$$

$$\begin{aligned} V_1 &= -x_3y_2 + x_4y_2 + x_2y_3 - x_4y_3 - x_2y_4 + x_3y_4 \\ &= (x_4 - x_2)(y_4 - y_3) - (x_4 - x_3)(y_4 - y_2) \\ &= (x_4 - x_2)(y_2 - y_3) - (x_2 - x_3)(y_4 - y_2) \\ &= (x_4 - x_3)(y_2 - y_3) - (x_2 - x_3)(y_4 - y_3) \end{aligned} \tag{16}$$

As can be seen from above, a constraint of $V_0 > 0$ or $V_1 < 0$ is essentially constraining the relationship between the slopes between the three points (points 1,2, and 3 for V_0, and points 2,3, and 4 for V_1).

In order for the entire CBS to convex, it does not suffice for the signs of V_0 and V_1 to be both positive. We must additionally ensure that equation 11 does not have a root between 0 and 1. Again, we use a monotonic transformation$z = t/(1 - t)$, in which case the relevant part (the part that modulates the sign) of equation 11 becomes the following: $V_1z^2 + (q + 2r)z + V_0$. Then, the two roots of this formula is as follows:

$$\frac{-(q+2r) \pm \sqrt{(q+2r)^2 - 4V_0V_1}}{2V_1} \tag{17}$$

If the determinant is negative, it means the following constraint is true:

$$-2\sqrt{V_0V_1} < q + 2r < 2\sqrt{V_0V_1} \tag{18}$$

If the determinant is non-negative, it means the converse is true:

$$q + 2r \leq -2\sqrt{V_0V_1} \quad or \quad 2\sqrt{V_0V_1} \leq q + 2r \tag{19}$$

Since $V_1 > 0$, we know that the denominator of equation 16 is positive and hence the numerator must be negative. If the left part of inequality 18 is true, it follows that both roots of equation 16 is positive. Hence only the right side of inequality 18 can hold. Combining the constraints, we have the following for a fully convex curve:

$$V_0 > 0, \qquad V_1 > 0, \qquad -2\sqrt{V_0V_1} < q + 2r \tag{20}$$

While this is a sufficient condition of a fully convex curve, it is difficult to utilize because the underlying variables are intertwined. If we use the monotonicity constraint from the compulsory condition (9), we can simplify this (20) further.

First, we prove that $V_0 > 0, V_1 > 0, 0 < q + 2r$ is a sufficient condition for convexity using proof by contradiction. Let's assume that there is a fully convex curve that satisfies the following constraints: $V_0 > 0, V_1 > 0, -2\sqrt{V_0V_1} < q + 2r \leq 0$.

If$x_2 < x_3$, it implies $x_1 < x_2 < x_3 < x_4$ by (9), and the following conditions hold:

$$V_0 > 0 \Leftrightarrow (x_2 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_1) > 0 \Leftrightarrow \frac{y_3 - y_1}{x_3 - x_1} > \frac{y_2 - y_1}{x_2 - x_1} \tag{21}$$

$$V_1 > 0 \Leftrightarrow (x_4 - x_2)(y_4 - y_3) - (x_4 - x_3)(y_4 - y_2) > 0 \Leftrightarrow \frac{y_4 - y_3}{x_4 - x_3} > \frac{y_4 - y_2}{x_4 - x_2} \tag{22}$$

$$q + 2r \leq 0 \Leftrightarrow (x_2 - x_1)(y_4 - y_3) - (x_4 - x_3)(y_2 - y_1) \leq 0 \Leftrightarrow \frac{y_4 - y_3}{x_4 - x_3} \leq \frac{y_2 - y_1}{x_2 - x_1} \tag{23}$$

Also, by the inequality of arithmetic and geometric means, we also see that if $-2\sqrt{V_0V_1} < q + 2r$ holds, $-(V_0 + V_1) < q + 2r$ also holds, giving us the following inequality:

$$-(V_0 + V_1) < q + 2r \Leftrightarrow (x_3 - x_1)(y_4 - y_2) - (x_4 - x_2)(y_3 - y_1) > 0 \Leftrightarrow \frac{y_4 - y_2}{x_4 - x_2} > \frac{y_3 - y_1}{x_3 - x_1} \tag{24}$$

Combination of inequality 21, 22, and 24 give us the following inequality:

$$\frac{y_4 - y_3}{x_4 - x_3} > \frac{y_4 - y_2}{x_4 - x_2} > \frac{y_3 - y_1}{x_3 - x_1} > \frac{y_2 - y_1}{x_2 - x_1} \tag{25}$$

However, this is directly contradicted by inequality 23. Hence $x_2 < x_3$ cannot hold.

If $x_2 \geq x_3$, we expand the condition of $-2\sqrt{V_0V_1} < q + 2r \leq 0$:

$$-2\sqrt{\big((x_2-x_3)(y_2-y_1)-(x_2-x_1)(y_2-y_3)\big)\big((x_4-x_3)(y_2-y_3)-(x_2-x_3)(y_4-y_3)\big)}<$$
$$(x_2-x_1)(y_4-y_3)-(x_4-x_3)(y_2-y_1)\le 0 \tag{26}$$

From the compulsory condition on equation 9, we have $\sqrt{(x_4-x_3)(x_2-x_1)}>x_2-x_3$. We can multiply both sides by$\sqrt{(x_4-x_3)(x_2-x_1)}$, which gives us $(x_4-x_3)(x_2-x_1)>(x_2-x_3)\sqrt{(x_4-x_3)(x_2-x_1)}$. Since both sides of inequality 26 are negative, we can divide the left side with a smaller positive number and the right side with a larger positive number and still maintain the inequality. Hence we divide the left side with $(x_2-x_3)\sqrt{(x_4-x_3)(x_2-x_1)}$ and the right side with$(x_4-x_3)(x_2-x_1)$. This gives us the following:

$$-2\sqrt{\left(\frac{y_2-y_1}{x_2-x_1}-\frac{y_2-y_3}{x_2-x_3}\right)\left(\frac{y_2-y_3}{x_2-x_3}-\frac{y_4-y_3}{x_4-x_3}\right)}<\frac{y_4-y_3}{x_4-x_3}-\frac{y_2-y_1}{x_2-x_1} \tag{27}$$

Applying the inequality of arithmetic and geometric mean on the left side, we have the following:

$$-\left(\frac{y_2-y_1}{x_2-x_1}-\frac{y_2-y_3}{x_2-x_3}+\frac{y_2-y_3}{x_2-x_3}-\frac{y_4-y_3}{x_4-x_3}\right)<\frac{y_4-y_3}{x_4-x_3}-\frac{y_2-y_1}{x_2-x_1} \tag{28}$$

However, the left-side equals the right side and the inequality provides a contradiction. Hence whether $x_2<x_3$ or $x_2\ge x_3$, our initial assumption of $V_0>0, V_1>0, -2\sqrt{V_0V_1}<q+2r\le 0$ does not hold. Concordantly, we have $V_0>0, V_1>0, 0<q+2r$.

Now if we assume $x_3\le x_2$, these conditions give us contradiction:

$$V_0>0, V_1>0 \Leftrightarrow \frac{y_4-y_3}{x_4-x_3}<\frac{y_3-y_2}{x_3-x_2}<\frac{y_2-y_1}{x_2-x_1} \tag{29}$$

$$q+2r>0 \Leftrightarrow \frac{y_4-y_3}{x_4-x_3}>\frac{y_2-y_1}{x_2-x_1} \tag{30}$$

Hence, $x_2<x_3$ and the convexity condition can be simplified as the following:

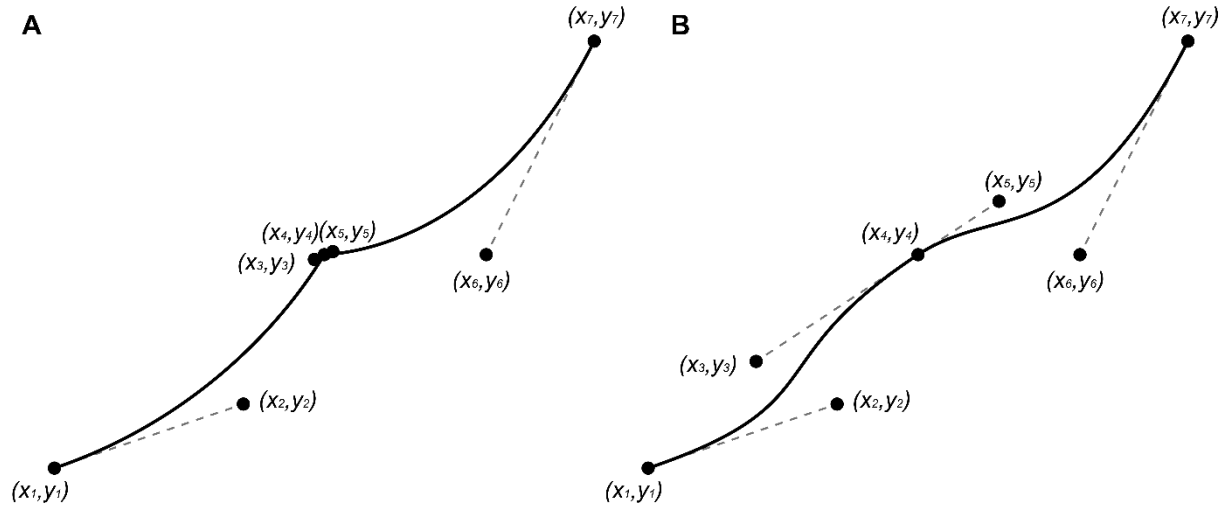
$$\frac{y_2-y_1}{x_2-x_1}<\frac{y_3-y_2}{x_3-x_2}<\frac{y_4-y_3}{x_4-x_3} \tag{31}$$

3. Smoothness Conditions of multi-piece CBS

So far, all the constraints have been worked out with regards to a single piece of CBS. However, in order to model more complex functions, one would need to chain multiple pieces of CBS functions together. In order to guarantee a smooth transition between the two chained CBS functions, there are some constraints that one should impose.

One constraint should be fairly obvious: we need to ensure that the derivative of CBS is continuous at the joining point of two CBS functions. This is straightforward as the line between the anchor point and its control point marks the local derivative. Hence, one should just ensure that the two control points of an anchor point is on the same line.

The second constraint may not always be necessary, but it can be useful. If the control handles (i.e., the distance between anchor point and its control points) becomes very short, there is potential for a kink in that location (see figure 3). Therefore, it can be useful to constrain the minimal distance between control points and their anchor point.



Supplement Figure 6. Two-piece CBS. *Panel A on the left shows the kink that forms when the control points become too close to the anchor point at the joining point of two CBS functions. Panel B shows a smooth transition between two pieces of CBS functions thanks to the appropriate distance between control points and anchor point.*

The table below summarizes all the constraints.

Compulsory	$x_1 < x_2 < x_4, \quad x_1 < x_3 < x_4$	
Slope	Monotonically increasing	$y_1 < y_2 < y_4, \quad y_1 < y_3 < y_4$
	Monotonically decreasing	$y_1 > y_2 > y_4, \quad y_1 > y_3 > y_4$
Curvature	Convex	$\frac{y_2 - y_1}{x_2 - x_1} < \frac{y_3 - y_2}{x_3 - x_2} < \frac{y_4 - y_3}{x_4 - x_3}$
	Concave	$\frac{y_2 - y_1}{x_2 - x_1} > \frac{y_3 - y_2}{x_3 - x_2} > \frac{y_4 - y_3}{x_4 - x_3}$

Smoothness	For a *n* – piece CBS function, $\frac{y_i - y_{i-1}}{x_i - x_{i-1}} = \frac{y_{i+1} - y_i}{x_{i+1} - x_i}, \quad \forall i = 3j + 1, \quad j \in \{1,2, \dots, n-1\}.$ And, $m^2 < (x_i - x_{i-1})^2 + (y_i - y_{i-1})^2, \quad m^2 < (x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2$ For a small number *m* (*m* = 0.1 in this paper).

4. Parameter Recovery of Prospect Theory

Here we show a small simulation that illustrates the confound in parameter recovery of prospect theory models. We generate choice data using a 2-parameter Prelec function (Prelec, 1998) and show that the recovered parameters show a clear correlation between the elevation of the probability weighting function and the curvature of the value function.

For simplicity, we generate choice data using only one parameter combination: $\gamma = 0.5$, $\delta = 1$, $\alpha = 0.8$. The δ parameter governs the elevation of the probability weighting function and α governs the curvature of the value function. We employed stimuli from three different datasets. One is the Kable et al. (2017) stimuli of 120 binary choices between a simple gamble and a certain amount. Second is the stimuli from Erev et al. (2002) that has 200 binary choices between two simple gambles. Third is the stimuli from Stott (2006) that has 90 binary choices between two two-outcome gambles. The second and the third dataset's parameter recovery capabilities have been examined in Broomell & Bhatia (2014), but only with regards to a 1-parameter Prelec function instead of the 2-parameter that we are using.

Choice probabilities were generated using a logit specification:

$$p(choice = 1) = \left(1 + \exp\left(-b(U_1 - U_2)\right)\right)^{-1} \quad (32)$$

We used cumulative prospect theory model with 2-parameter Prelec function for calculation of utilities. For stimuli from Kable et al. (2017) and Erev et al. (2002), this was the following:

$$U = f(p) \cdot A^{\alpha} \quad (33)$$

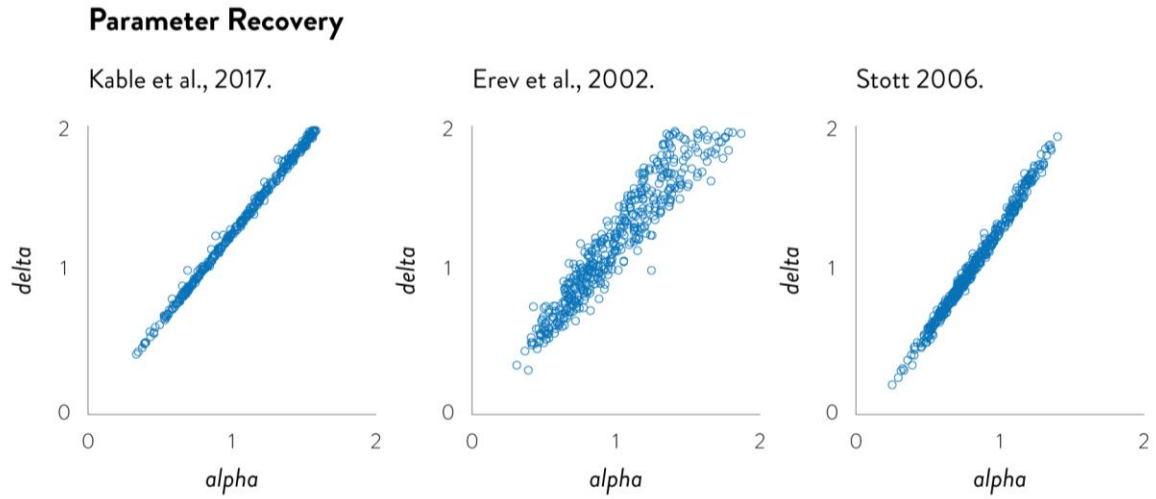
Where the *p* is the probability of winning the simple gamble and *A* is the amount. In the case of stimuli from Stott (2006), the specification for a two-outcome gamble of winning a large amount (A) with probability *p* and a smaller amount (SA) with probability *1-p* becomes the following:

$$U = f(p) \cdot A^{\alpha} + \left(1 - f(p)\right) \cdot SA^{\alpha}. \quad (34)$$

In both cases $f(p) = \exp[-\delta(-\ln p)^{\gamma}]$. Since the scaling factor *b* in eq. 32 also controls the choice stochasticity, we set it such that on average, each simulated choice set would differ from the true preference in 3 out of 100 choices ($b = 7.1$ for Kable et al. (2017), $b = 3.3$ for Erev et al. (2002), and $b = 0.063$ for Stott (2006)). Given the choice probabilities, we simulated 500 choice sets for each of the three stimuli set and then fitted each of the 1,500 choice sets using the same model as the generating model. Then we examined the pattern between the fitted parameters.

There was a clear, strong correlation between the elevation parameter δ and the curvature parameter α. The three panels from supplement Figure 4 shows the scatterplot of the recovered parameters. All three stimuli sets resulted in a near-linear correlation between the two parameters.

Furthermore, the range of fitted values were very broad, suggesting that these two parameters are hardly identifiable. While the true value of alpha was 0.8, its fitted values ranged anywhere between 0.2 and 1.5, resulting in a variety of value functions from extremely diminishing marginal returns to extremely increasing marginal returns. The fitted elevation parameter delta was also widely varying between 0.3 and 2, resulting in diverse probability weighting functions that show from overall overweighting of probabilities to overall underweighting of probabilities. In over 91% of all simulations, the fitted parameter combination resulted in higher log-likelihood than the true parameter combinations, which suggest that these results are not due to failed convergence of maximum likelihood estimation.



Supplement Figure 4. Parameter recovery results. *The three panel shows the scatterplot of recovered parameters alpha and delta for each of the three stimuli sets.*