# Flexible 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

# Flexible Utility Function Approximation via Cubic Bezier Splines

Abstract: In intertemporal and risky choice decisions, parametric utility models are widely used for predicting choice and measuring individuals' impulsivity and risk aversion. However, parametric utility models cannot describe data deviating from their assumed functional form. We propose a novel method using Cubic Bezier Splines (CBS) to flexibly model smooth and monotonic utility functions that can be fit to any dataset. CBS shows higher descriptive and predictive accuracy over extant parametric models and can identify common yet novel patterns of behavior previously unaccounted for. Furthermore, CBS provides measures of impulsivity and risk aversion that do not depend on parametric model assumptions.

Keywords: flexible modeling, heterogeneity, intertemporal choice, risky choice, generalized utility functions

Intertemporal choices (ITCs) and risky choices (RCs) are heavily studied across many disciplines. ITCs are decisions regarding outcomes that occur at different times: for example, deciding between spending money now versus saving and investing that money for later, smoking now versus having better health later, or whether to pay an additional price for expedited shipping in order to receive a package earlier. RCs are decisions made regarding outcomes that occur probabilistically: buying lottery tickets, investing in stock markets, gambling, etc. ITCs and RCs are studied both in basic and applied research. In basic research, researchers are interested in how people make ITCs or RCs and have generated many different proposals for the utility functions that underlie these choices. In applied research, researchers are often interested in how individual differences in ITC and RC relate to real world behaviors such as pathological gambling, smoking, susceptibility to mental illness, drug and alcohol abuse, education level and financial status (Alessi & Petry, 2003; Anderson & Mellor, 2008; Brañas-Garza, Georgantzís, & Guillén, 2007; Kirby, Petry, Nancy, & Bickel, Warren, 1999; Krain et al., 2008; Lejuez, Aklin, Bornovalova, & Moolchan, 2005; Lejuez et al., 2003; Lempert, Steinglass, Pinto, Kable, & Simpson, 2019; Schepis, McFetridge, Chaplin, Sinha, & Krishnan-Sarin, 2011; Shamosh & Gray, 2008)(Alessi & Petry, 2003; Anderson & Mellor, 2008; Brañas-Garza et al., 2007; Kirby et al., 1999; Krain et al., 2008; Lejuez et al., 2005, 2003; Schepis et al., 2011; Shamosh & Gray, 2008).

ITC and RC research, be it basic or applied, relies heavily on parametric utility models (**Table 1**). Parametric models have two benefits that make them popular. First, parametric utility models can distill complex patterns of behavior into one or two interpretable parameters. For example, the discount rate parameter in ITC models represent the rate at which the value of future time options decline; the risk-aversion parameter in RC models (often substituted by the

value function curvature parameter) represent the deviation of utilities from risk-neutral expected value. These interpretable parameters are especially useful in applied research that seeks to correlate these measures with other variables such as health or intelligence. Of course, in order to obtain these measures, one must fit the models to data, which brings us to the second benefit of parametric models: simplicity in usage. Unlike non-parametric approaches that require very carefully planned experimental designs or a large number of data-points (e.g., Abdellaoui, 2003; Myerson, Green, & Warusawitharana, 2006; Wakker & Deneffe, 1996), parametric utility models can be nested inside logit or probit choice models and fit to any dataset using simple procedures such as maximum likelihood estimation (MLE). Hence, parametric utility models provide a simple and interpretable method for describing behavior in ITC and RC.

However, established parametric models in ITC and RC have difficulty accounting for an important phenomenon: the heterogeneity and non-regularity in utility function shapes. While parametric utility models can allow for some heterogeneity in their free parameters that alter the shape of the function, recent evidence shows that different people behave according to different utility model forms altogether and that there is no 'one correct model' that can describe everyone's behavior equally well (Bruhin, Fehr-Duda, & Epper, 2009; Cavagnaro, Aranovich, McClure, Pitt, & Myung, 2016; Franck, Koffarnus, House, & Bickel, 2015; Myerson et al., 2006). Adding to this concern, even if a multitude of models are considered (e.g., by using mixture models), there is no guarantee that the data comes from one of the considered utility functions. This evidence for heterogeneity is troubling for the large body of research in ITC and RC that relies on assumed parametric utility models to make descriptions and predictions of behavior and to make inferences from empirical data.

In sum, while parametric models are useful in their simplicity and interpretability, their

assumptions can be questionable at the individual level due to heterogeneous utility function forms that cannot be accounted for by simply allowing for heterogeneity in parameters. Given this, we believe that a flexible semi-parametric approach (as in Cubic Bezier Splines described here) to utility modeling can provide many benefits over parametric utility models.

First, a flexible utility model can better describe and predict ITC and RC by avoiding model misspecification. Model misspecification errors are likely if one assumes the same parametric utility function for everyone even if the free parameters are allowed to vary across people. Hence, instead of assuming a single parametric model, we can approximate a much greater class of models; by avoiding model misspecification, a flexible approximation of individually unique functions can lead to increased descriptive and predictive powers in ITC and RC.

Second, a flexible approach can identify novel patterns of behavior. Existing parametric models were developed to account for established behavioral phenomena at the population level; therefore, these model assume that patterns of behavior fall within a certain range, thereby being unable to identify novel patterns that fall outside this range. For example, in ITC, parametric models typically assume constant or decreasing discount rates over time. As an example, the commonly employed exponential model assumes a constant multiplicative discount rate as can be seen from its discounting function: $f(D) = e^{-kD} = (e^{-k})^D = \delta^D$. More generally, the discount rate at a given delay $D^*$ can be calculated as $h(D^*) = \ln(-\ln(f(D^*))/D^*)$, which is a constant in the case of exponential function: $\ln(-\ln(e^{-kD^*}) / D^*) = \ln(k)$. All other common models, as shown in **Fig. 1A**, show decreasing discount rates over time. RC models too, also only allow for certain behavioral patterns; RC parametric models typically assume that people cannot alternate between risk-averse and risk-seeking behavior more than once across

probabilities. For example, the expected utility model ($U = A^{\alpha} \cdot p$ in a gamble of winning $A$ with probability $p$ or nothing otherwise) assumes that a person is either risk-averse or risk-seeking throughout all probabilities depending on whether the value function curvature parameter ($\alpha$) is below 1 or above 1, respectively. More generally, if we convert the RC models into a discounting form of $U = A \cdot f(p)$, we can measure the degree of risk-aversion at a given probability $p^*$ by $q(p^*) = \ln(f(p^*)/p^*)$, which is the log odds of subjective to objective probabilities. As shown in **Fig. 1B**, expected utility theory and hyperbolic models assume that people are risk-averse or risk-seeking throughout all probabilities, while prospect theory models and generalized hyperbolic models assume that people's behavior can 'switch' at most once from risk-seeking to risk-aversion (or vice versa) as probabilities increase (indicated by the change of sign in $q(p^*)$). In both ITC and RC, then, existing parametric models cannot identify consistent patterns of behavior that go outside of the range specified by these model's assumptions. However, a flexible approach that makes no assumptions of specific behavioral patterns can detect novel behaviors unaccounted for by existing models, as we will demonstrate empirically.

Third, a flexible approach can provide interpretable measures of impulsivity and risk-aversion without assuming a parametric utility model. These measures of impulsivity and risk-aversion are of wide interest in applied ITC and RC research. Currently, measures of impulsivity and risk-aversion are obtained by fitting the free parameters in a parametric utility model (e.g., the discount rate in ITC models, value function curvature in RC models). Consequently, researchers often face the concern that their findings may be dependent on their choice of parametric model and the assumptions that come with it. To allay these concerns, researchers often perform the same analysis multiple times using different utility models to show the robustness of their results (e.g., Ballard & Knutson, 2009; Kable & Glimcher, 2007). However,

not only is this a burden on researchers whose research is not focused on the 'correct' form of utility function, but it is also an imperfect solution as there always could be another model to consider. A more effective solution is to use a flexible utility model that does not make rigid assumptions about the form of the utility function, so that the measure of impulsivity and risk aversion is based more on the empirical data than on the parametric model assumptions (Myerson et al., 2006).

In this paper, we propose a flexible utility modeling tool based on cubic Bezier splines and show empirically that 1) it provides better description and prediction of our ITC and RC data as measured by in-sample and out-of-sample metrics, 2) it can be used to find novel patterns of behavior in both ITC and RC, including some individuals who show increasing discount rates over time in ITC and some who show switch multiple times between risk-aversion and risk-seeking across probabilities in RC, and 3) it provides measures of impulsivity and risk aversion without parametric model assumptions. We also provide statistical packages in MATLAB and R to be used for future research (see below).

## *CUBIC BEZIER SPLINES MODEL SPECIFICATION*

What are cubic Bezier splines (CBS; de Casteljau, 1963), and why did we choose CBS over alternative methods to provide flexible utility modeling? In this section, we briefly explain what qualities we were looking for in a flexible method, how that ruled out other possible methods, and how CBS can be used as a choice model in ITC and RC.

There were three primary criteria we wanted in a flexible utility model: ease of use, interpretability, and normative constraints. Ease of use was of primary importance as the currently existing methods for flexible utility models could not be used on most datasets. A

common non-parametric utility modeling procedure is to use interpolations to construct utility functions; however, these procedures require specialized datasets such as adaptive choices or choice sets constructed according to a specific mathematical formula (Abdellaoui, 2003; Myerson et al., 2006; Wakker & Deneffe, 1996). Unlike such methods, our goal was to provide a general likelihood-based method utility model that could be embedded in a logit or probit choice model so that it can be applied to any choice dataset just like parametric models.

Our second criterion was interpretability: rather than 'black-box' models such as neural nets, gaussian processes, regression trees, etc., we wanted to have an interpretable structure to the utility model and its output. Hence, we opted for the discounted utility form of $U = A \cdot f(X)$; in ITC, this would be $U = A \cdot f(D)$ where amount (A) is discounted as a function of delay (D), and in RC, this would be $U = A \cdot f(p)$ where amount (A) is discounted as a function of probability ($p$). The discounting form has several benefits: 1) most ITC models are already in discounting form, 2) most RC models could be analytically converted to discounting forms, 3) discounting forms are easily identifiable through choice, and 4) the discounting form allows a measure of impulsivity and risk aversion to be solely contained in one fitted function, which makes interpretation of the utility function easy (see **supplemental materials** for details on RC model conversion and identifiability of prospect theory forms). Embedding this discounted utility function inside a binary logit choice model gives us the following specification:

$$\log\left(\frac{p(choice_t = 1)}{p(choice_t = 2)}\right) = \sigma(U_{1t} - U_{2t}), \qquad U_{jt} = A_{jt} \cdot f\left(X_{jt}\right), \qquad j = 1{,}2 \qquad (1)$$

where $\sigma$ is a free parameter that determines the relationship between the scale of the utilities ($U_{1t}$, $U_{2t}$) and choice, and $X_{jt}$ is either delay or probability, depending on the task. Hence, the key question came down to this: how to flexibly approximate $f\left(X_{jt}\right)$?

Our last criterion, in regards to specifying $f(X_{jt})$, was an ability to incorporate two normative constriants: smoothness and monotonicity. Given a continuously smooth input variable such as delay or probability, it made normative sense that the output variable of utility was also continuously smooth. In ITC, it makes normative sense for utility to decline monotonically as a function of delay while in RC, to increase monotonically as a function of probability. The two normative constraints of smoothness and monotonicity are already implicit in almost all of the existing parametric utility models and can serve as important priors that combat over-flexibility. Hence the goal was to estimate a smooth, monotonic univariate transformation of $f(X_{jt})$. However, the monotonicity constraint makes the use of several methods difficult. Polynomial or fourier basis regressions, while continuously smooth, control the flexibility of the curve by changing the order of the equation, which unfortunately also changes the order of the derivative and complicates the constraining problem (see **supplemental materials** for discussion on B-splines). Hence, we found that by chaining multiple pieces of cubic-order Bezier splines, each of them separately monotonically constrained, we can approximate $f(X)$ in a smooth, monotonic manner, without requiring speicalized datasets.

Piecewise connected CBS are already widely used in graphics software, fonts, and interpolations, but have seen limited use as function approximators compared to other types of splines. This is because while most splines are defined in the form of $y = f(x)$, where the $y$ coordinate is expressed as a function of $x$, CBS's functional form is much more general: both the $x$ and $y$ coordinates are independently expressed as functions of a third variable $t$. A single piece of CBS is defined by four points ($P_{0x}$, $P_{0y}$), ($P_{1x}$, $P_{1y}$), ($P_{2x}$, $P_{2y}$), ($P_{3x}$, $P_{3y}$) (**Fig. 2A**). The coordinates of these four points become the parameters of the CBS as the x and y-coordinates of the spline are controlled independently by two separate cubic functions.

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

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

These two functions can jointly be used to approximate the function $f(D)$ in ITC or $f(p)$ in RC by $y = n\left(m^{-1}(x)\right)$ as long as $x = m(t)$ and $y = n(t)$ are both monotonic functions of *t*. We found that the constraint for monotonicity is very simple: if the x and y coordinates of the two middle points (P1 and P2) stay between that of the end points (P0 and P3), the resulting CBS is monotonic (i.e., $P_{1x}, P_{2x} \in [P_{0x}, P_{3x}]$, and $P_{1y}, P_{2y} \in \left[P_{0y}, P_{3y}\right]$; see **supplemental materials** for proof). It is also important to note that the CBS's local derivative at the end point equals the slope of the line connecting the end point with its neighboring point (i.e., $\overline{P_2 P_3}$ in **Fig. 2**). Using this property, multiple pieces of CBS can be smoothly joined by equating the local derivative (i.e., ensuring that three points *P₂ P₃ P₄* are on the same line in **Fig. 2B**). **Fig. 3** shows the CBS parameters involved in modeling $f(D)$ and $f(p)$ in ITC and RC using either 1-piece or 2-pieces of CBS.

Because the CBS form ($y = n\left(m^{-1}(x)\right)$; **eq. 2 and 3**) cannot be succinctly expressed as $y = f(x)$, the likelihood function for the choice model using CBS also cannot be succinctly expressed. Instead, shown below is the general MLE steps (a pseudo-algorithm) used to fit a CBS-based choice model:

1. Start with some initial CBS points (**Fig. 3** shows relevant points for each case)
2. For all $X_{jt}$ (delay or probability), find $t_{jt}^*$ that satisfies $X_{jt} = m\left(t_{jt}^*\right)$ as given in **eq. 2**.
3. Then, calculate $U_{jt} = A_{jt} \cdot n(t_{jt}^*)$ as given in **eq. 3**
4. Use **eq. 1** to calculate the log-likelihood of all choices

5. Propose new parameters using gradient descent while maintaining constraints in **Fig. 3**.
6. Repeat step 2 through 6 until convergence

In this paper, the CBS functions were fit using MATLAB's optimization tool (fmincon). For this paper, we only entertain 1-piece and 2-piece CBS as they seemed sufficient in approximating the parametric utility models shown in **Table 1** (utility function recovery simulation results are shown in **supplemental materials**). The MATLAB code used to fit the data is available online as a package in a ready-to-use form (https://github.com/sangillee/CBSm), and the R version of the package can be downloaded from CRAN under the name 'CBSr' (currently going through review at CRAN), and represents another contribution of this research.

## *EMPIRICAL METHODS*

Using real ITC and RC data, we demonstrate the benefits of a CBS modeling approach. We utilize ITC and RC data collected in Kable et al. (2017). 166 participants completed binary choice tasks in ITC and RC and 128 of them returned after 10 weeks to perform the same task again in session 2. In each session, participants made 120 binary choices each in the ITC task and RC task. The choices in the ITC tasks 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 D days; D ~ [20 180], Y ~ [22,85]). The choices in the RC tasks were between a smaller certain monetary reward that was always \$20 and a larger probabilistic monetary reward (e.g., \$Y with probability $p$; $p$ ~ [.09 .98], Y ~ [21 85]). We treated session 1 and 2 as if they are separate participants and only included sessions with at least two or more of each choice types (i.e., at least two smaller reward choices and two larger reward choices in 120 trials), which ruled out 9 sessions for ITC and 4 sessions in RC. 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 (i.e., all smaller reward choices or all larger reward choices).

We empirically show the three benefits of CBS's flexible utility function. First, we compare the descriptive and predictive capabilities of CBS against other parametric models in **Table 1**. For all models (including CBS), we measured their in-sample and out-of-sample prediction accuracies and Tjur's D. Tjur's D (coefficient of discrimination) is the difference of the mean choice probabilities of each choice type. For example, a good model of ITC 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 between 0 (random model) and 1 (perfect model) and tells how well the two choice types are discriminated in out-of-sample predictions. Even if two models have the same hit rate accuracy, Tjur's D is higher for models that classify the trials with larger discrimination in choice probabilities. All models were fit using a logit choice model of the following form:

$$\log\left(\frac{p(choice_t = 1)}{p(choice_t = 2)}\right) = \sigma(U_{1t} - U_{2t}) \quad (4)$$

where $\sigma$ modeled the overall scale of the utility difference between the two options. The utilities of each options at each trial ($U_{1t}$, $U_{2t}$) were modeled according to the form shown on **Table 1**.

Second, we use CBS to demonstrate novel patterns of behavior. Using the out-of-sample prediction values of Tjur's D, we show that a portion of participants are better fit by CBS models because they exhibit behavior violating the aforementioned assumptions of extant parametric models. Specifically, we show that, in ITC, some participants exhibited increasing discount rate over time, and in RC, some participants exhibited multiple alternations between risk aversion and risk seeking across probabilities.

Third and finally, we use CBS to obtain measures of impulsivity and risk-aversion without assuming a parametric utility model. The CBS measures of impulsivity and risk-aversion can be obtained by measuring the area under the curve (AUC) of the fitted CBS function (see **supplemental materials** for analytic expression). Since CBS models ITC and RC utility in discounting form ($U = A \cdot f(X)$), the AUC of the discounting function $f(X)$ serves as a measure of how much the amount is discounted as a function of delay or probability. In previous research, AUC of discounting form utility functions has been proposed and used as a measure of impulsivity and risk-aversion in non-parametric utility estimation (Myerson et al., 2006). We show that the AUC of CBS fits can serve as stable, subject-specific measure of impulsivity and risk-aversion, just like the parameter estimates of extant parametric models, by testing cross-session consistency (i.e., correlation) of ITC and RC AUC.

## *EMPIRICAL RESULTS*

### *Increased descriptive & predictive power*

CBS showed higher in-sample and out-of-sample accuracies and Tjur's D than all of the tested parametric models (**Fig 4**). For both in-sample accuracy and Tjur's D, we found that 2-piece CBS function provides performance superior to all other methods in both ITC and RC, followed by 1-piece CBS. These in-sample results are somewhat expected given that models with more parameters are generally more likely to provide higher performance metrics. However, even in out-of-sample prediction CBS provides the highest accuracy and Tjur's D compared to all other parametric models in both ITC and RC. This clearly demonstrates that CBS is not simply providing a flexible function that overfits empirical data; rather its flexibility is important in capturing individual characteristics so as to increase descriptive and predictive

power. In ITC, out-of-sample accuracy was highest for the 1-piece CBS model followed by the 2-piece CBS model, while out-of-sample Tjur's D was highest for 2-piece CBS model followed by the 1-piece CBS model. This may suggest that while 1-piece CBS model may provide the highest hit rate accuracy, the 2-piece CBS model may be able to better separate the two choice types. In RC, both the out-of-sample accuracy and Tjur's D was highest for the 2-piece CBS model with substantial lead over the next runner-up 1-piece CBS model. This pattern may suggest that RC data may generally require more complex functions to adequately model behavior compared to ITC data as we describe below.

To provide further insight, we show eight example participants' data that span the variety of choice patterns in the data. Their choices and their model fits are shown in **Fig. 5** and **Fig. 6**. Both the fit and the choices are shown in relative amounts. The relative amount is the immediate amount divided by the larger amount; for example, in ITC, 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 of 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 two choice types (shown in circles and Xs). Panel A shows 4 participants' data whose highest LOOCV Tjur's D 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. Given larger datasets, CBS will likely match the parametric models in these participants. Panel B, on the other hand, shows 4 participants' data whose highest LOOCV Tjur's D 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 fits a rather unconventional, but flexible monotonic function that separates the two choice

types.

More specifically, participants who are best fit by CBS as shown in **Fig. 5** and **Fig. 6** seem to exhibit behaviors that cannot be accounted for by parametric models. **Fig. 5** shows that in ITC, several participants seem to exhibit a utility function that decreases sharply at certain delays. Such sharp decreases in utility indicate suddenly increasing impatience and discount rates, which cannot be accounted for by any of the parametric models we considered. **Fig. 6** shows four participants who are best fit by CBS seemingly due to their utility function having more complexity via multiple inflection points. Generally, it seems that the participants are risk-averse in low probabilities (as shown by the fitted curve being below the identity line), while being risk-seeking around $p = .5$, and being risk-averse again above .5. This pattern of multiple switches between risk-aversion and risk-seeking behavior deviates from the established parametric models which can only account for either overall risk-aversion or risk-seeking throughout all probabilities, or a one-time switch between risk-aversion and risk-seeking.

*Identifying novel patterns of behavior*

Using fitted CBS functions, we identified novel systematic patterns of behavior in both ITC and RC. In ITC, we found many participants have increasing daily discount rates, and hence were unaccounted for by the parametric models (**Fig. 7A, 7B**). When we grouped the fitted CBS functions based on the average daily change of discount rate, the best parametric model's LOOCV Tjur's D was as good as that of CBS models when participants had decreasing discount rates, which is the commonly assumed pattern. However, when the average daily discount rates were increasing, the CBS models significantly outperformed the best parametric models in LOOCV (**Fig. 7A**). When we examined the fitted CBS functions, we found that when the average daily change was negative, the median CBS function looked very similar to other ITC

parametric models; on the other hand, when the average daily discount rate was positive (i.e., increasing discount rate over time), we found that the median of the CBS functions became linear or even concave, neither of which could be accounted for by parametric ITC models.

In RC, we found that many participants switch multiple times between risk-aversion and risk-seeking as probabilities increase; a behavior that is unaccounted for by the parametric models as all parametric models of RC allow either no switching between risk-aversion and risk-seeking or a single switch point. Concordantly, CBS's LOOCV Tjur's D were significantly higher than the best parametric models' LOOCV Tjur's D for participants with 2 or more switches (**Fig. 7C**). This result suggests that participants exhibit potentially much more complex patterns of behavior than what parametric models have assumed. Interestingly, even in participants that do not switch between risk-aversion and risk-seeking, we found that CBS significantly outperforms other parametric models in LOOCV. **Fig. 7D** shows the median CBS fitted functions grouped by the number of switches between risk-averse and risk-seeking behavior (as seen by how many times the function crosses the identity line). We can see that when participants switch once, their average function resembles a typical prospect theory S-shaped function (albeit risk-averse in low probabilities). This simple form is likely captured well by most parametric utility models, thereby leading to similar predictive performance between parametric and CBS models. However, when participants' risk aversion switches twice or three times, the average function clearly cannot be captured by any of the parametric RC models. Furthermore, although the parametric utility models can account for non-switching behavior as well, we can see that the average function for non-switching behavior has some inflection points that cannot be captured by the parametric models (cf. **Fig. 1B**).

*Model-agnostic measures of impulsivity and risk aversion*

We found the CBS measures of impulsivity and risk-aversion to be highly correlated across the 2 sessions 10 weeks apart (**Fig. 8**). This result suggests that CBS's measure of impulsivity and risk-aversion can pick up stable individual traits that are often needed in applied ITC and RC research. Using the 2-piece CBS fits to the real choice data, we calculated, for each session, an overall measure of impulsivity and risk aversion by calculating the AUC of the fitted CBS function. The cross-session Pearson correlations of the AUCs were very high at $r = 0.79$ ($p < .001$) for ITC and $r = 0.60$ ($p < .001$) for RC. These measures were comparable to the cross-session consistencies of extant parametric models' impulsivity and risk aversion measures; the hyperbolic model's discount rate (logk) had cross-session correlation of $r = 0.80$, and EUT's risk-aversion measure (log $\alpha$) had cross-session correlation of $r = 0.65$. This shows that CBS's measure of impulsivity and risk aversion can provide a stable individual-specific measure of overall impatience or risk aversion without assuming a fully parametric model.

## *DISCUSSION*

Cubic Bezier Splines are a promising flexible method that can approximate individual utility functions without fully parametric assumptions. By relaxing the parametric assumptions, it can help researchers better account for heterogeneous utility functions in ITC and RC. Here we introduced how CBS can be used to parsimoniously approximate univariate functions, especially given monotonicity constraints. Using CBS, we have shown that by estimating a general class of utility models of the form $U = A \cdot f(X)$, we can approximate a wide class of models that include most of the often-used parametric models. Such properties allowed us to demonstrate three benefits of CBS method over extant parametric model approaches in ITC and RC.

First, CBS can provide individually tailored utility functions, which led to improved descriptive (in-sample) and predictive (out-of-sample) capabilities. In datasets with

heterogeneous utility functions, having an individually tailored utility function allows researchers to circumvent potential model misspecifications. Using CBS approximations provided a stronger defense against model misspecifications than entertaining a multitude of models as empirical data may not be describable by any known parametric models. We have shown, using a large empirical choice dataset, that CBS provides, in both in-sample and out-of-sample metrics, superior performance in description and prediction of choices compared to all of the often-used parametric utility models that we tested.

Second, CBS can be used to detect novel patterns of behavior that violate extant models' assumptions. In the current paper, we have identified two novel patterns of behavior from ITC and RC data. In ITC, we found that there are participants who exhibit increasing discount rates and therefore cannot be accounted for by the currently established parametric models of ITC. Such participants exhibited concave utility functions which may be indicative of a heuristic (e.g., deciding not to wait after a certain delay). In RC, we found that there are participants who alternate between risk-aversion and risk-seeking multiple times within the probability range of [0 1]. Such complex patterns of behavior could not be described by the established parametric models of RC which assume at most one switch between risk-aversion and risk-seeking behavior. Having found these patterns, future studies in ITC and RC may be able to identify new ways of clustering the said patterns to identify participants who may use different sets of psychological processes when making decisions (e.g., Reeck, Wall, & Johnson, 2017).

Third, CBS provides measures of impatience and risk aversion that do not depend on a specific parametric utility model. Given the heterogeneity of utility functions in choice data, there has always been a need to characterize individual's overall behavior without having to rely on a specific model (Myerson et al., 2006). The area under the curve (AUC) of the estimated

CBS function serves as an overall measure of impulsivity or risk aversion that is robust to model misspecifications even in the face of heterogeneous data. Furthermore, unlike the previous nonparametric approaches (e.g., Abdellaoui, 2003; Myerson, Green, & Warusawitharana, 2006; Wakker & Deneffe, 1996), CBS can be estimated from any choice dataset without large or specially structured datasets.

CBS also has the potential to aid other research questions. First, it can aid the study of choice stochasticity by more accurately dissociating between model misspecification and choice noise. Goodness-of-fit measures for parametric utility functions do not provide good assessments of choice noise because one cannot distinguish whether the data is stochastically noisy or if the utility model is simply misspecified. Previous research has focused on the monotonicity of utility functions to make a theoretical distinction between model misspecification and genuine noise (Johnson & Bickel, 2008). Since CBS models that we present here have only the general normative assumption of monotonicity, the noise estimates from CBS only includes the stochasticity that cannot be explained with a monotonic utility function. Future research may seek to correlate choice stochasticity with other measures such as impulsivity, risk-aversion, age, education, and/or IQ.

Second, CBS can provide more accurate estimates of latent utilities, which will also aid current efforts to relate such utilities to other behavioral and neural measures (Levy & Glimcher, 2012; Venkatraman, Payne, & Huettel, 2014). For example, numerous studies have examined drift-diffusion and similar models that can incorporate response time data into choice data (Busemeyer & Townsend, 1993; Clithero, 2018; Dai & Busemeyer, 2014; Forstmann, Ratcliff, & Wagenmakers, 2016; Ratcliff, Smith, Brown, & McKoon, 2016). By using utility estimates that can describe participants' choices better than traditional parametric utility estimates,

development and validation of these models can be improved. Also, decision neuroscience research often requires estimates of utilities that can be used to search for correlates of valuation in the brain (e.g., Kable & Glimcher, 2007; Knutson, 2005). These efforts can also benefit from more refined estimates of utility that better predicts participants' choices.

Despite these substantial benefits, it is important to note that there are some drawbacks of flexible approaches like CBS. CBS, or at least the current version, is best used on datasets that have reasonable coverage over a range of values, as there's no 'default' shape that CBS tends towards in absence of data. In future research, CBS's extrapolation capabilities can be enhanced under a Bayesian framework by using priors toward a commonly used utility function (e.g., hyperbolic, or EUT) such that CBS can default to more simple forms in the absence of data, but take on a more complex form given sufficient data. Alternatively, future studies may develop adaptive questioning schemes that are based on CBS estimations to provide an efficient flexible 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 flexible models to explore many topics not only related to economic decision-making, but also other cognitive, affective, and social behaviors whose models have latent variables. We hope that across many areas of human behavior, the behavioral patterns and heterogeneity that went unnoticed under formal parametric assumptions can now easily be brought to surface and studied.

Acknowledgements

This research was supported by grants from the NIMH (R01-MH098899), the NSF (1533623), and the National Cancer Institute (R01-CA170297, R35-CA197461). We would like to thank Fred Feinberg, Longxiu Tian, and Zach Winston for their helpful insights and discussions on this project.

*TABLES*

| Intertemporal Choice Models | Form Name | Utility Function | Approx. by CBS |
|---|---|---|---|
| Samuelson, 1937 | Exponential (E) | $U = A \cdot \exp(-kD)$ | Y |
| Mazur, 1987 | Hyperbolic (H) | $U = A \cdot (1 + kD)^{-1}$ | Y |
| Green, Fry, & Myerson, 1994 | Generalized Hyperbolic (Gh) | $U = A \cdot\ (1 + kD)^{-s}$ | Y |
| Roelofsma, 1996 | Log Time (Lt) | $U = A \cdot D^{-k}$ | Y |
| Laibson, 1997 | Quasi-hyperbolic (Q) | $U = A \cdot \beta \exp(-kD)$ | Y |
| McClure, et al., 2007 | Double Exponential (De) | $U = A \cdot (we^{-aD} + (1 - w)e^{-bD})$ | Y |

| Risky Choice Models | Form Name | Utility Function | Approx. by CBS |
|---|---|---|---|
| Von Neumann & Morgenstern, 1945 | Expected Utility Theory (Eut) | $U = A \cdot p^{1/\alpha}$ | *Y |
| Rachlin et al., 1991 | Hyperbolic (H) | $U = A \cdot \left(1 + h\left(\frac{1-p}{p}\right)\right)^{-1}$ | *Y |
| Goldstein & Einhorn, 1987 | GE-weight Prospect T. (Ge) | $U = A \cdot \left(\frac{\delta p^{\gamma}}{\delta p^{\gamma} + (1-p)^{\gamma}}\right)^{1/\alpha}$ | *Y |
| Tversky & Kahneman, 1992 | TK-weight Prospect T. (T) | $U = A \cdot \left(\frac{p^{\gamma}}{(p^{\gamma} + (1-p)^{\gamma})^{\frac{1}{\gamma}}}\right)^{1/\alpha}$ | *Y |
| Prelec, 1998 | Prelec-weight Prospect T. (P) | $U = A \cdot (\exp(-\delta(-\ln p)^{\gamma}))^{1/\alpha}$ | *Y |
| Green & Myerson, 2004 | Generalized Hyperbolic (Gh) | $U = A \cdot \left(1 + h\left(\frac{1-p}{p}\right)\right)^{-s}$ | *Y |
| Markowitz, 1959 | Risk-Return (R) | $U = A \cdot p - b \cdot Var$ | N |
| Slovic & Lichtenstein, 1968 | Attribute (A) | $U = \beta_0 + \beta_1 A + \beta_2 p$ | N |
| Weber et al., 2004 | Coefficient of Variation (C) | $U =\ A \cdot p - b \cdot CV$ | N |

**Table 1. Survey of commonly used ITC and RC models.** Each row shows, from left to right, the reference of the parametric model, the name of the form (with short abbreviation), the model specification, and whether the model can be approximated by the CBS function of the form in this paper. Across all ITC models, utility is expressed as a product of A, the amount of the delayed outcome, and f(D), which is a function of the delay. In RC models, A is the amount of the risky outcome, p is the probability of winning that outcome. We only show here the model forms for a simple gamble in which there is a probability p of winning A and probability 1-p of winning nothing. *The RC models marked with an asterisk are written in an analytically converted form that allows to be approximated by CBS in the form of $U = A \cdot f(p)$ (see **supplemental materials section A** for the conversion proof).

*FIGURES*

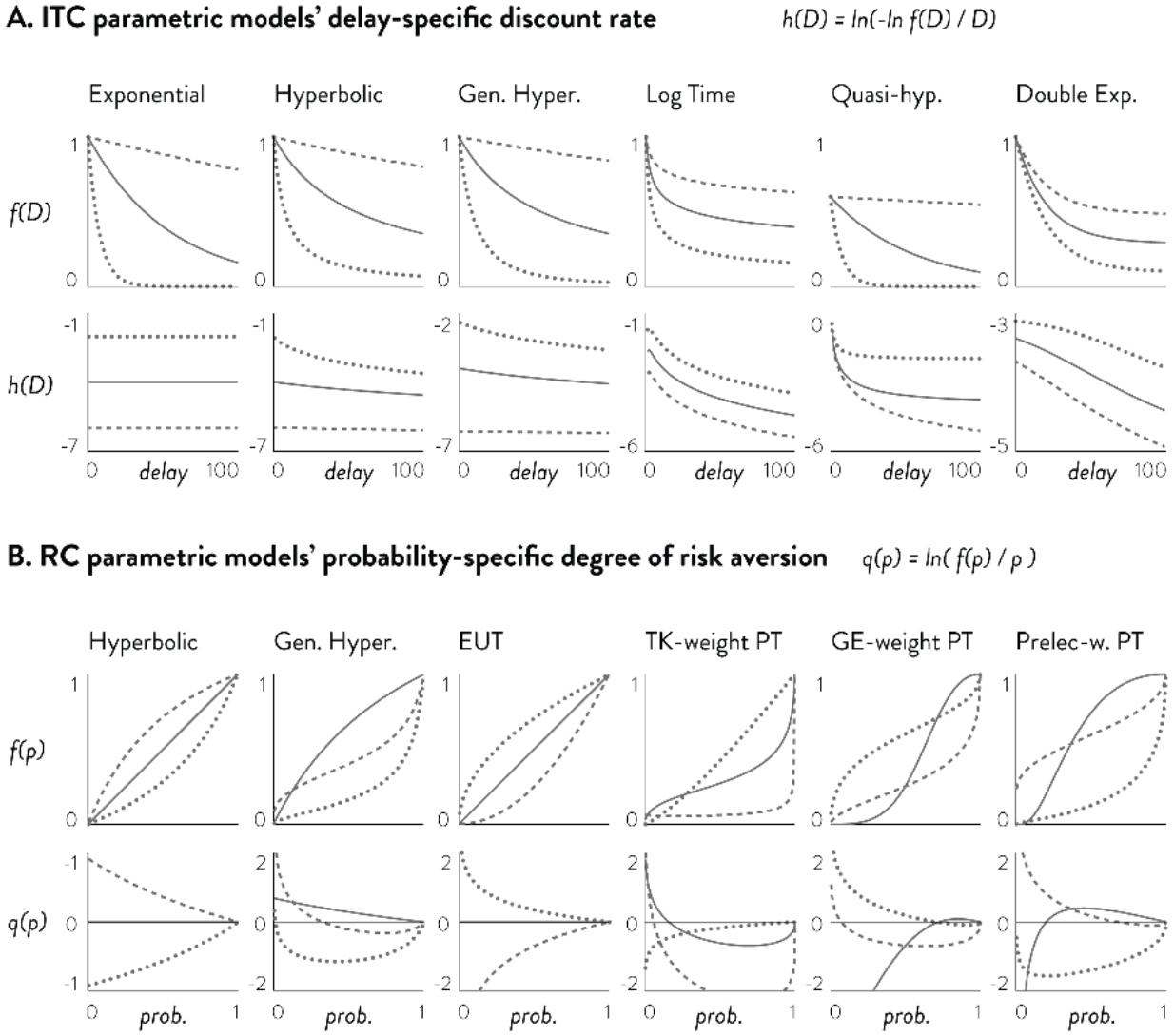



**Figure 1. Parametric models' pattern of behavior.** Each parametric model assumes a particular behavioral pattern. Shown in panel A is the delay-specific discount rate of ITC models in Table 1. All parametric models of ITC in consideration show either a constant (exponential) or decreasing delay-specific discount rates. Shown in panel B is the probability-specific degree of risk aversion, which is the log of the ratio between objective and subjective probabilities. A measure above 0 would indicate over-appreciation of probabilities and hence risk-seeking, while a measure below 0 would indicate risk-aversion. All parametric models of RC in consideration assume a behavioral pattern that switches between risk-aversion and risk-seeking at most once. In other words, probability-specific measure of risky choice for all RC parametric models can cross 0 (risk-neutral point) at most once.

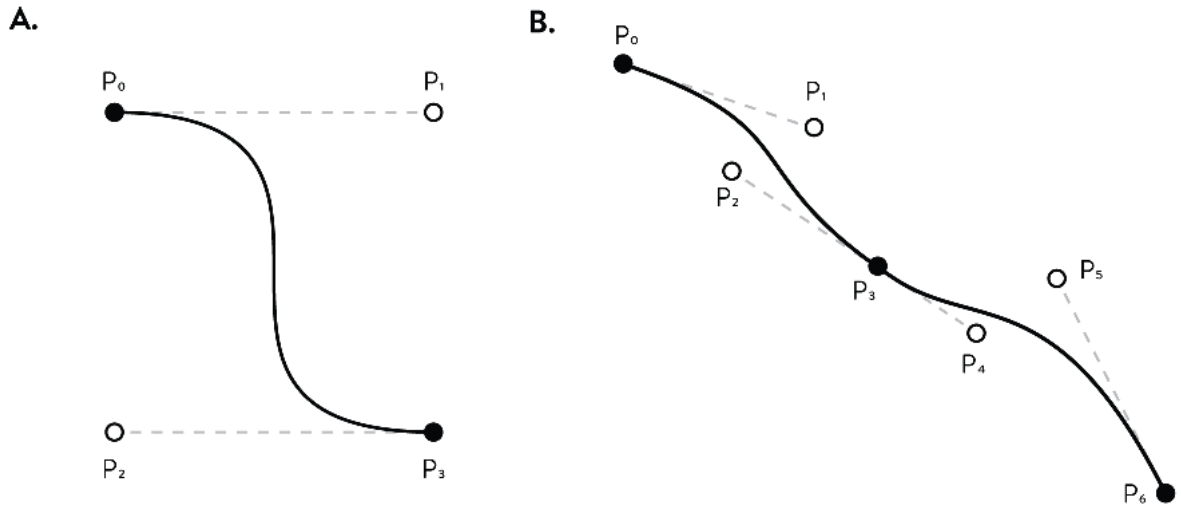



**Figure 2. 1-piece and 2-piece CBS (A and B, respectively).** Example 1-piece CBS is shown in panel A, and 2-piece CBS is shown in panel B. While each piece requires 4 points, because adjoining points overlap, 2-piece CBS requires 7 points.

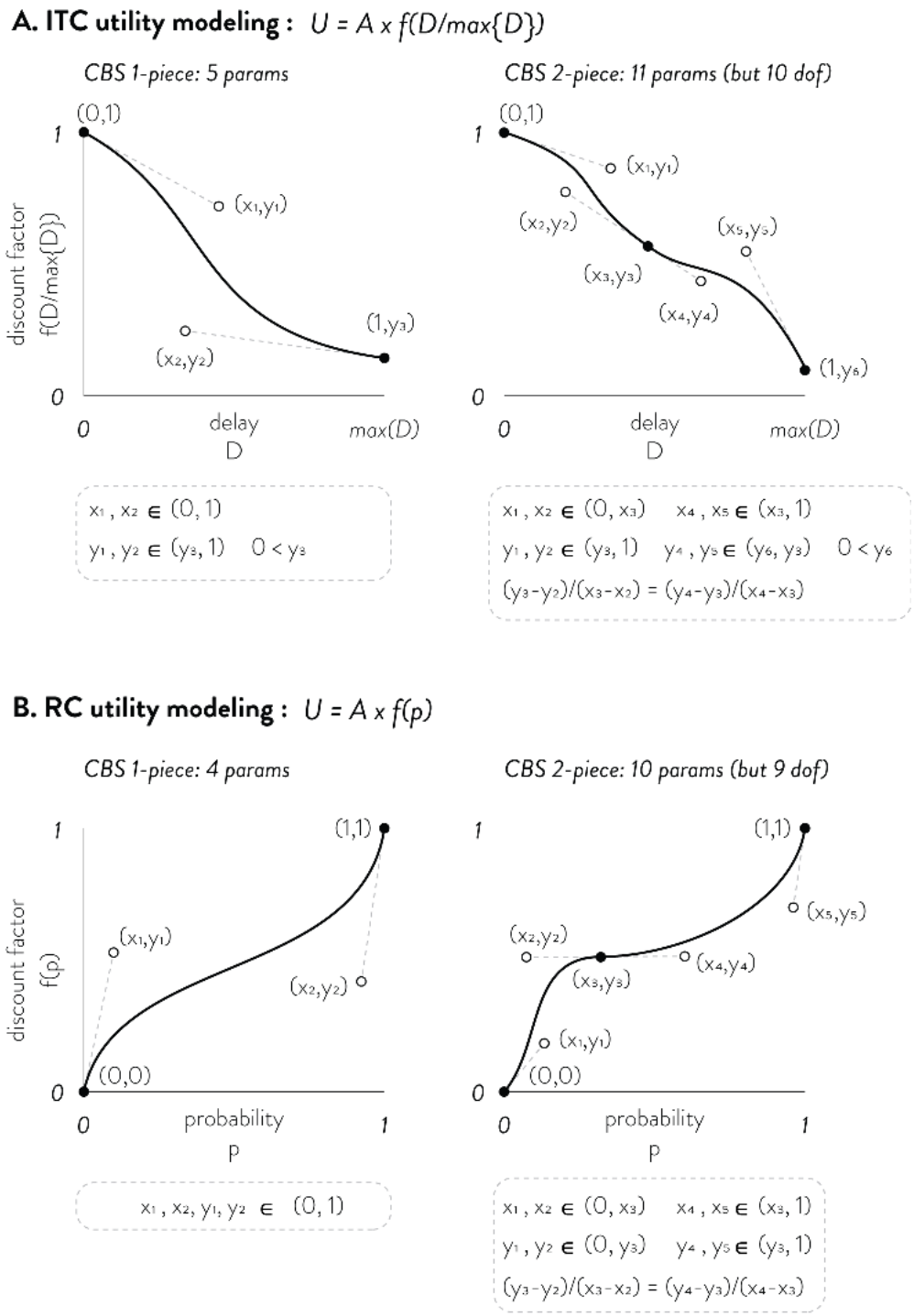



**Figure 3. Model specification of ITC (A) and RC (B) using 1-piece (left) and 2-piece CBS (right).** Panel A above shows how CBS is used to flexibly model the delay discounting function and panel B shows how CBS models the monotonically increasing probability distortions. In both ITC and RC, the coordinates of the points are free parameters that are estimated. The parameter constraints are shown below each panel in dotted boxes. In the case of 2-piece CBS, there is one less degrees of freedom than number of parameters due to the necessity of $(x_2,y_2)$, $(x_3,y_3)$, and $(x_4,y_4)$ being on the same line.

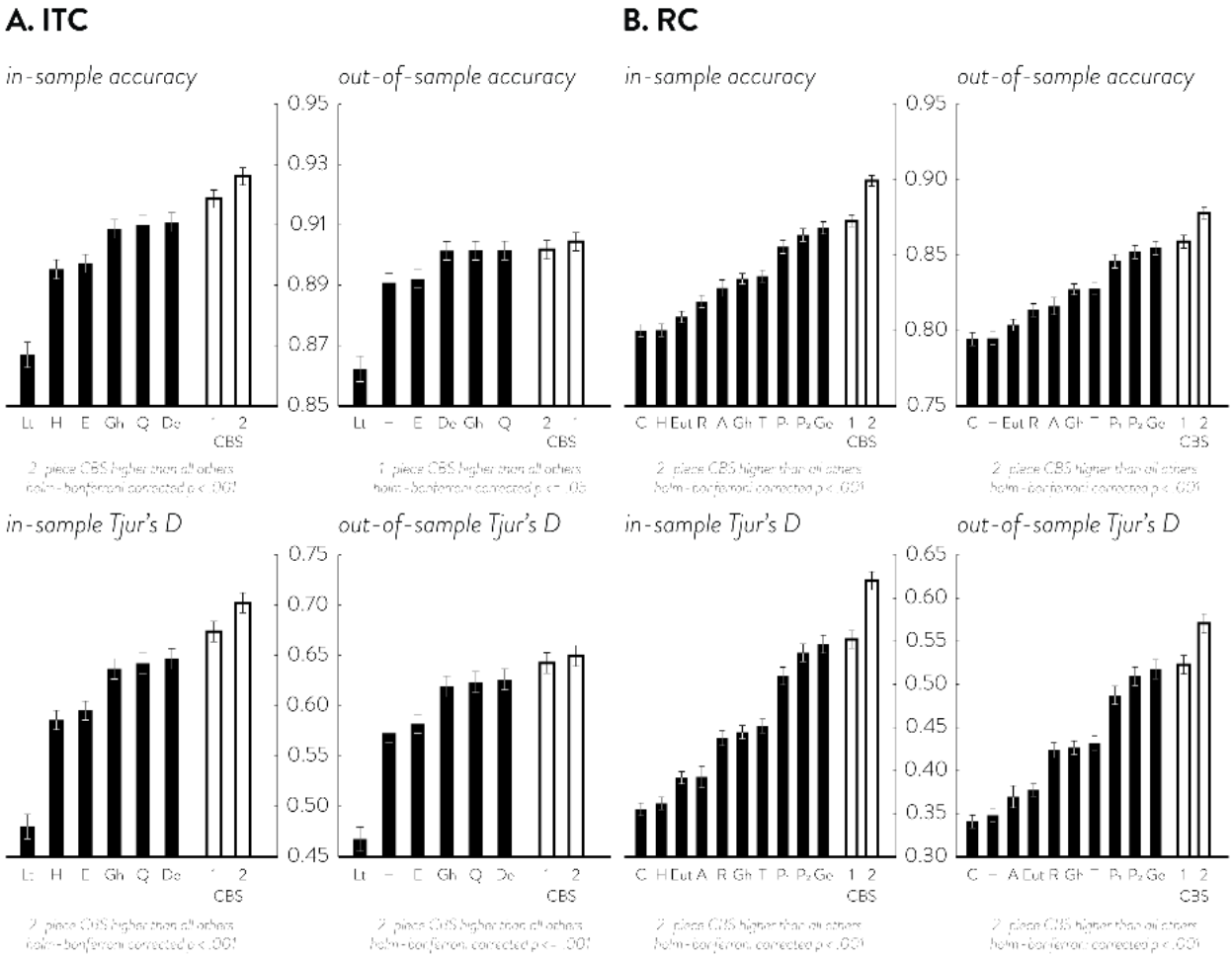



**Figure 4. In-sample and out-of-sample prediction performance in ITC (A, left), and RC (B right).** In ITC, 6 parametric models and 2 CBS models were assessed; in RC, 10 parametric models and 2 CBS models were assessed. In both in-sample and out-of-sample, each model's accuracy (top row) and Tjur's D (bottom row) were assessed. The error bars represent the standard error of the mean.

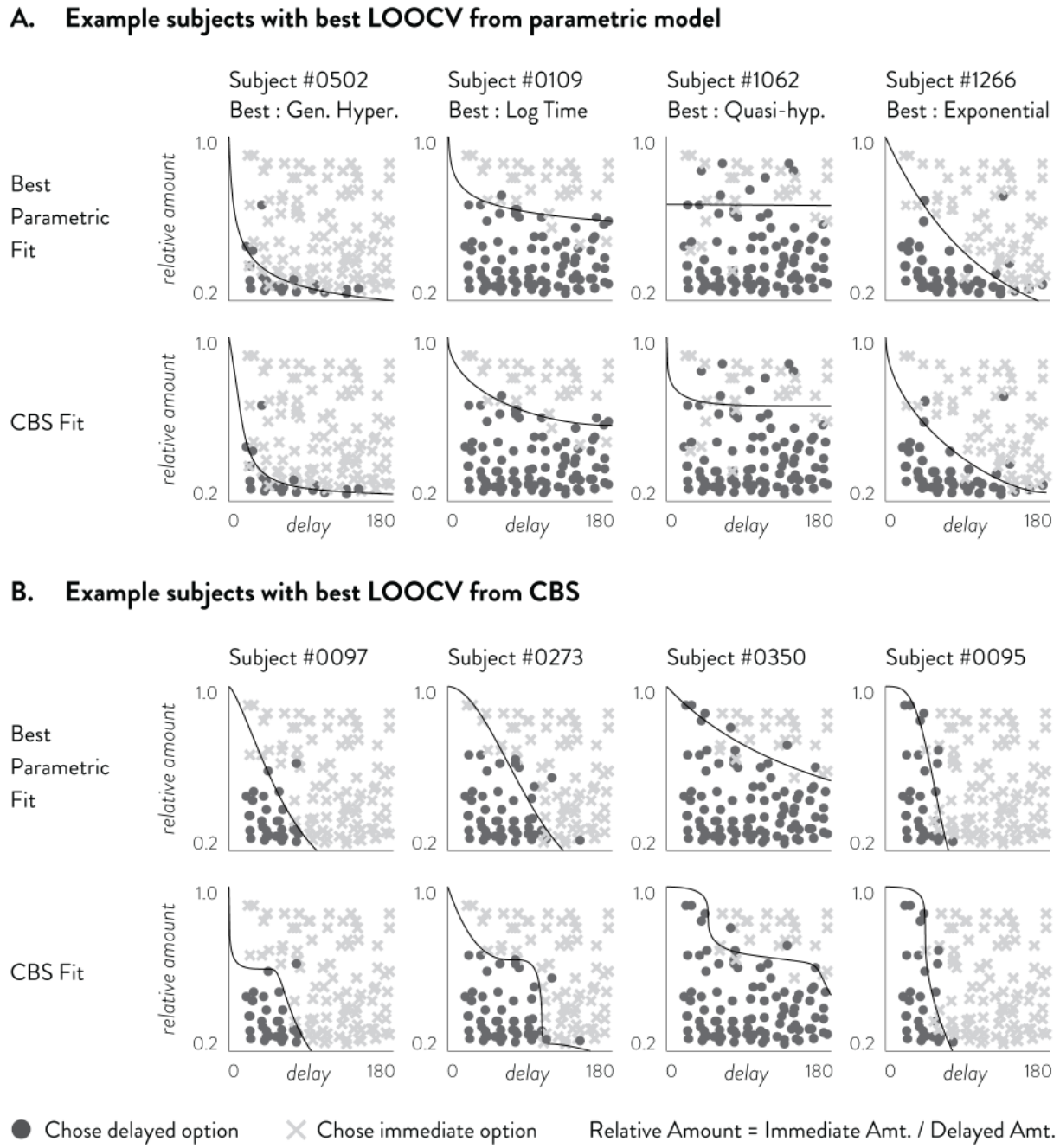



**Figure 5. Plots of eight example participants' choices, their best parametric fits and their best CBS fits as determined by LOOCV.** Panel A shows 4 participants whose highest LOOCV Tjur's D came from parametric models and panel B shows 4 participants whose highest LOOCV Tjur's D came from CBS. In each panel, the top row shows the best parametric model (by LOOCV) and the bottom row shows the CBS fit. Panel A participants were selected such that the diverse parametric forms can be shown; panel B participants were selected to show variety of CBS fits that did not conform to parametric forms.

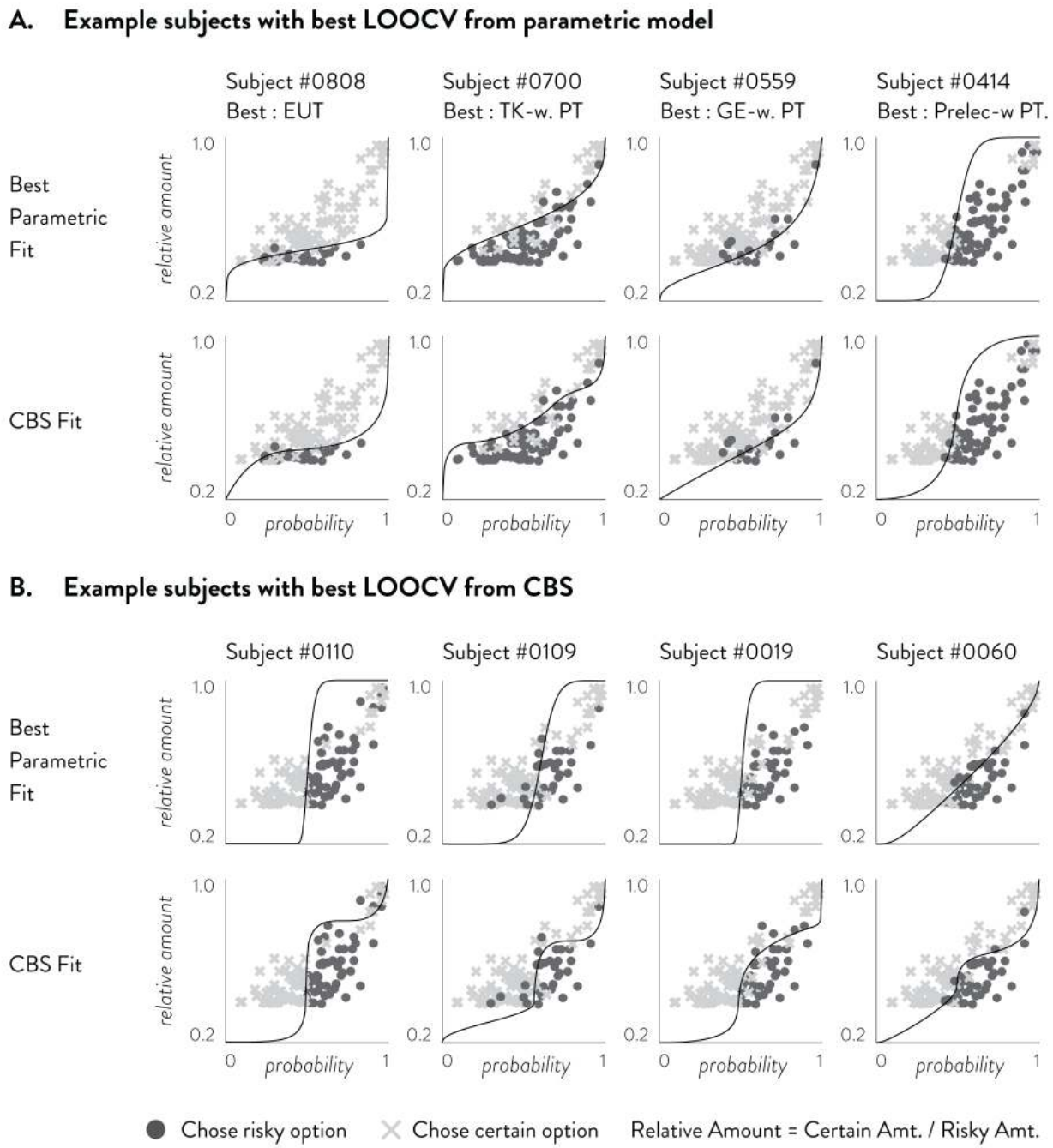



**Figure 6. Plots of eight example participants' choices, their best parametric fits and their best CBS fits as determined by LOOCV.** Panel A shows 4 participants whose highest LOOCV Tjur's D came from parametric models and panel B shows 4 participants whose highest LOOCV Tjur's D came from CBS. In each panel, the top row shows the best parametric model (by LOOCV) and the bottom row shows the CBS fit. Panel A participants were selected such that the diverse parametric forms can be shown; panel B participants were selected to show variety of CBS fits that did not conform to parametric forms.

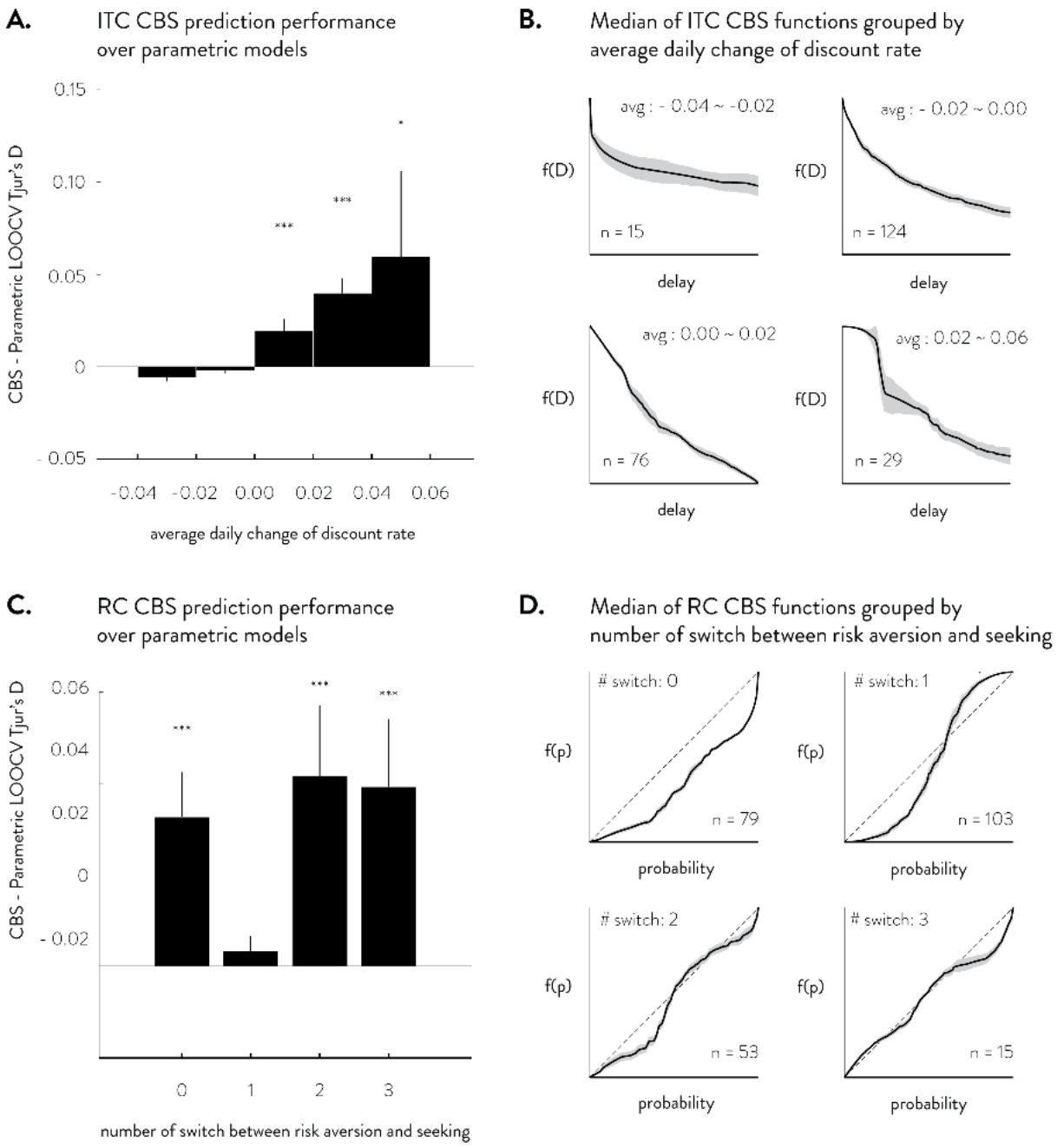



**Figure 7. Deviation from common parametric forms.** Panel A shows CBS prediction performance minus the maximum of parametric models' prediction performance. CBS shows increasingly better predictions as the average daily change in discount rate becomes positive. Panel B shows the average fitted CBS functions for ITC grouped by the average daily change of discount rate. The solid line is the median function, with gray shade showing the standard errors. Panel C shows that, in RC, CBS provides better predictions to participants who do not alternate between risk aversion and risk seeking or alternate more than one time. Panel D shows the average fitted CBS functions for RC grouped by the number of switches between risk-aversion and risk-seeking behavior. * t-test against 0, $p < .05$. *** $p < .001$.

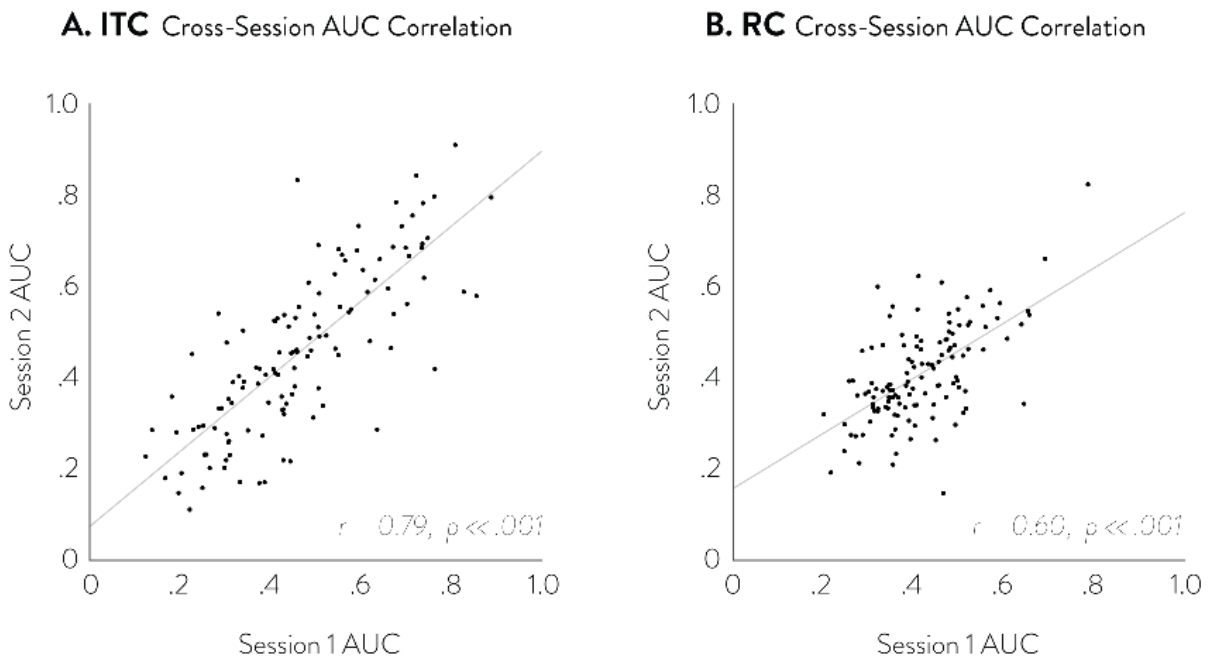



**Figure 8. Cross-session correlation of overall measure of delay discounting (left) and risk aversion (right) as measured by Area Under the Curve (AUC) of CBS.** The abscissa marks the AUC measure of each participant in session 1 and the ordinate marks the AUC measure of each participant in session 2. The cross-session Pearson correlation measure of AUC was 0.79 for ITC and 0.60 for RC, both with p-values less than .001.

*REFERENCES*

Abdellaoui, M. (2003). Parameter-Free Elicitation of Utility and Probability Weighting Functions. *Management Science*, *46*(11), 1497–1512.

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.

Ballard, K., & Knutson, B. (2009). Dissociable neural representations of future reward magnitude and delay during temporal discounting. *NeuroImage*, *45*(1), 143–150.

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.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*. https://doi.org/10.1037/0033-295X.100.3.432

Cavagnaro, D. R., Aranovich, G. J., McClure, S. M., Pitt, M. A., & Myung, J. I. (2016). On the functional form of temporal discounting: An optimized adaptive test. *Journal of Risk and Uncertainty*. https://doi.org/10.1007/s11166-016-9242-y

Clithero, J. A. (2018). Improving out-of-sample predictions using response times and a model of the decision process. *Journal of Economic Behavior and Organization*. https://doi.org/10.1016/j.jebo.2018.02.007

Dai, J., & Busemeyer, J. R. (2014). A probabilistic, dynamic, and attribute-wise model of

intertemporal choice. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/a0035976

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

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E.-J. (2016). Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions. *Annual Review of Psychology*. https://doi.org/10.1146/annurev-psych-122414-033645

Franck, C. T., Koffarnus, M. N., House, L. L., & Bickel, W. K. (2015). Accurate characterization of delay discounting: A multiple model approach using approximate bayesian model selection and a unified discounting measure. *Journal of the Experimental Analysis of Behavior*. https://doi.org/10.1002/jeab.128

Goldstein, W. M., & Einhorn, H. J. (1987). Expression Theory and the Preference Reversal Phenomena. *Psychological Review*, *94*(2), 236–254.

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. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychological Bulletin*. https://doi.org/10.1037/0033-2909.130.5.769

Johnson, M. W., & Bickel, W. K. (2008). An Algorithm for Identifying Nonsystematic Delay-Discounting Data. *Experimental and Clinical Psychopharmacology*. https://doi.org/10.1037/1064-1297.16.3.264

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.

Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value druing

intertemporal choice. *Nature Neuroscience*, *10*(12), 1625–1633.

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.

Knutson, B. (2005). Distributed Neural Representation of Expected Value. *Journal of Neuroscience*, *25*(19), 4806–4812. https://doi.org/10.1523/JNEUROSCI.0642-05.2005

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.

Lempert, K. M., Steinglass, J. E., Pinto, A., Kable, J. W., & Simpson, H. B. (2019). Can delay discounting deliver on the promise of RDoC? *Psychological Medicine*. https://doi.org/10.1017/S0033291718001770

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: efficient diversification of investments. New Haven,

CT: Cowles Foundation.

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. *In: Commons, M.L., Mazur, J.E., Nevin, J.A., Rachlin, H*.

McClure, S. M., Ericson, K. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2007). Time Discounting for Primary Rewards. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.4246-06.2007

Myerson, J., Green, L., & Warusawitharana, M. (2006). Area under the curve as a measure of discounting. *Journal of the Experimental Analysis of Behavior*, *76*(2), 235–243.

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.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2016.01.007

Reeck, C., Wall, D., & Johnson, E. J. (2017). Search predicts and changes patience in intertemporal choice. *Proceedings of the National Academy of Sciences*, *114*(45), 11890–11895.

Roelofsma, P. H. M. P. (1996). Modelling intertemporal choices: An anomaly approach. *Acta Psychologica*.

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.

Slovic, P., & Lichtenstein, S. (1968). Relative importance of probabilities and payoffs in risk taking. *Journal of Experimental Psychology*. https://doi.org/10.1037/h0026468

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.

Von Neumann, J., & Morgenstern, O. (1945). Theory of games and economic behavior. *Bull. Amer. Math. Soc*, *51*(7), 498–504.

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.

*Supplemental Materials*

## A. Transformation of various RC models

Unlike ITC models that are mostly in the form of $U = A \cdot f(X)$, several RC models have either the prospect theory form of $U = f(A) \cdot g(p)$, or the probability discounting form of $U = A \cdot g(\theta)$ where $\theta = (1 - p) / p$. Hence it is theoretically easier for CBS to nest a variety of forms if we use the general form of $U = f(A) \cdot g(p)$. However, there are several reasons why we decided to use CBS to approximate the general structure of $U = A \cdot f(p)$.

First, other forms can be analytically converted into the form of $U = A \cdot f(p)$. Prospect theory forms can be analytically converted into a simple form in the case of a simple gamble, which is very widely used. Consider a simple gamble and its certainty equivalent: $CE^{\alpha} = A^{\alpha} \cdot wp$. We can de-exponentiate both sides to achieve $CE = A \cdot wp^{1/\alpha}$. Since $wp$ is being raised to an exponent, it does not alter its values at 0 and 1 which still remain 0 and 1 after exponentiation. Hence this can be estimated generally using the form $U = A \cdot f(p)$ with CBS. Discounting functions in the form of $U = A \cdot g(\theta)$ can also be converted by plugging in $\theta = (1 - p) / p$. For example, in the case of the generalized hyperbolic function $U = A / (1+h\theta)^{s}$, it can be converted into $U = A \cdot (1-h+h/p)^{-s}$. As can be seen from the formula, when $p = 0$, the right term goes to 0 and when $p = 1$, the right term goes to 1.

The second reason is that the form of $U = A \cdot f(p)$ 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 converted 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. Any point at which $f(p)$ is above the identity line is indicative of risk-seeking behavior, while any point below the identity line is indicative of risk-averse behavior.

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 $U = f(A) \cdot g(p)$; if one's $f(A)$ is generally high, one can lower the $g(p)$ to compensate to achieve the same utility. 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 appendix E that clearly shows the confound between $f(A)$ and $g(p)$.

## B. Monotonic B-splines and relationship to CBS

Basis-splines (more commonly referred to as B-splines), which CBS is a special case of, have been used in scenarios with monotonicity constraints (Brezger & Steiner, 2008; Leitenstorfer & Tutz, 2007); but the monotonicity constraint of B-splines has only been worked out for evenly spaced knots, which does not allow the data to determine the appropriate position of the knots. In this paper, we provide a method for allowing the data to control the position of the knots, thereby providing a more flexible approach with a fewer number of parameters (knots). This is done by connecting multiple small B-splines, each of which only has four knots, and can analytically be constrained for monotonicity even with free positioning of the knots. Here we use Bernstein basis function for the splines which makes them Bezier splines.

**C. Compulsory and Slope Conditions of CBS**

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:

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

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 that the final constraint is a non-linear inequality that entangles the 4 coordinates simultaneously, thereby reducing their independence. 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 $y$.

**D. 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:

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

$$
\begin{aligned}
&= (x_2 - x_3)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_3) \\
&= (x_2 - x_3)(y_2 - y_1) - (x_2 - x_1)(y_2 - y_3) \\
V_1 = &-x_3y_2 + x_4y_2 + x_2y_3 - x_4y_3 - x_2y_4 + x_3y_4 \\
&= (x_4 - x_2)(y_4 - y_3) - (x_4 - x_3)(y_4 - y_2) \\
&= (x_4 - x_2)(y_2 - y_3) - (x_2 - x_3)(y_4 - y_2) \\
&= (x_4 - x_3)(y_2 - y_3) - (x_2 - x_3)(y_4 - y_3)
\end{aligned} \tag{16}
$$

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: $V_1 z^2 + (q + 2r)z + V_0$. 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:

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

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

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

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

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

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

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

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 $-2\sqrt{V_0 V_1} < q + 2r \leq 0$:

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

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:

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

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

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

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

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

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

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

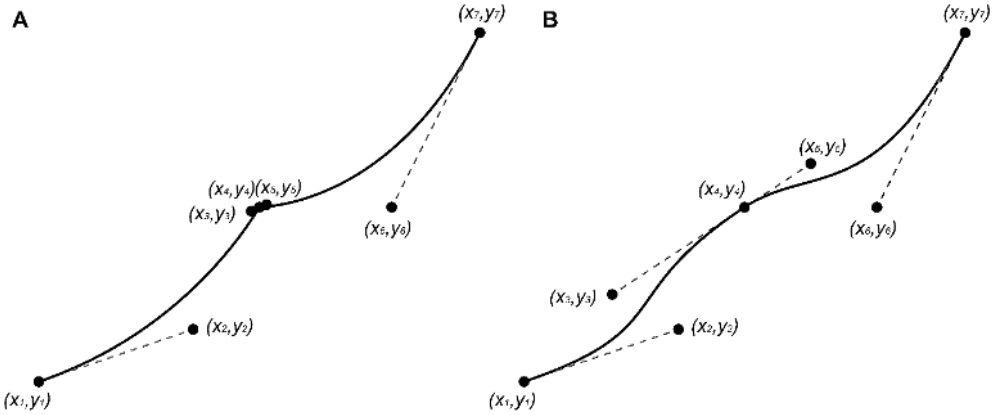
Hence, $x_2 < x_3$ and the convexity condition can be simplified as the following:

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

**E. 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.

First, the derivative of CBS must be 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. Second, 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 (**Supplemental Fig 1**). Therefore, it can be useful to constrain the minimal distance between control points and their anchor point.




***Supplemental Figure 1. 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,<br>$\frac{y_i - y_{i-1}}{x_i - x_{i-1}} = \frac{y_{i+1} - y_i}{x_{i+1} - x_i}, \quad \forall i = 3j + 1, \quad j \in \{1,2,\dots,n-1\}.$<br>And,<br>$m^2 < (x_i - x_{i-1})^2 + (y_i - y_{i-1})^2, \quad m^2 < (x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2$<br>For a small number $m$ ($m$ = 0.1 in this paper). | |

**F. Parameter Recovery of Prospect Theory**

Prospect theory forms can be extremely hard to estimate from simple gambles. This is not because the estimation doesn't converge, but rather due to collinearity that exists between two parameters: value function exponent and probability weighting function elevation parameter. Here we show in three ways why this would be the case.

Intuitively, both parameters control the same behavior. Consider a fair gamble where one option is \$20 for sure and the other is 50% chance of winning \$40. A risk-averse agent would generally choose the \$20. However, this could be either because the agent has an underappreciated probability for 50% and feel that the chance is not good, or because the agent feels that \$40 is not worth twice as much as \$20. The former would be in the domain of the probability weighting function while the latter would be in the domain of value function. Hence it is clear that both functions can explain the same behavior to a certain degree.

Mathematically, the confound is the clearest in Prelec's 2-parameter probability weighting function (1998). This is by no means a drawback of Prelec's function as other functions also share the same confound; rather if anything, it would be the advantage of Prelec's function for clarifying this relationship. Consider an equivalence in utility between a certain smaller monetary option (SA) and a larger risky monetary option (LA) with probability *p*. Then, the equation would be of the following:

$$SA^{\alpha} = LA^{\alpha} \cdot e^{-\delta(-lnp)^{r}} \tag{32}$$

Note that by de-exponentiating both sides of the equation, one can achieve the following:

$$SA = LA \cdot e^{-\frac{\delta}{\alpha}(-lnp)^{r}} \tag{33}$$

This is simply a model where the value exponent is 1 and the Prelec probability weighting function has an elevation parameter $\delta/\alpha$. So, the two parameters are entirely colinear in describing the relationship between two utilities.

Simulation-wise, this confound results in the estimated parameters varying wildly by little noise. 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: $r = 0.5$, $\delta = 1$, $\alpha = 0.8$. 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:

$$p(choice = 1) = \left(1 + \exp\left(-b(U_1 - U_2)\right)\right)^{-1} \tag{34}$$

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:

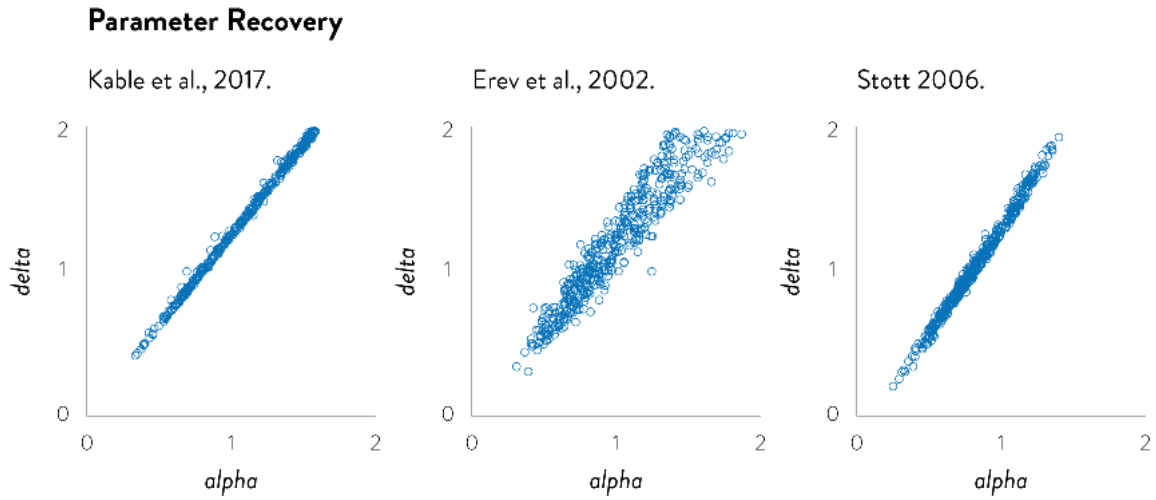
$$U = f(p) \cdot A^{\alpha} \tag{35}$$

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:

$$U = f(p) \cdot A^{\alpha} + \big(1 - f(p)\big) \cdot SA^{\alpha}. \tag{36}$$

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 $\delta$ and the curvature parameter $\alpha$. The three panels from **Supplemental Fig 2** 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.




***Supplemental Figure 2. Parameter recovery results.*** *The three panel shows the scatterplot of recovered parameters alpha and delta for each of the three stimuli sets.*

**G. Simulated Utility Function Recovery Using CBS**

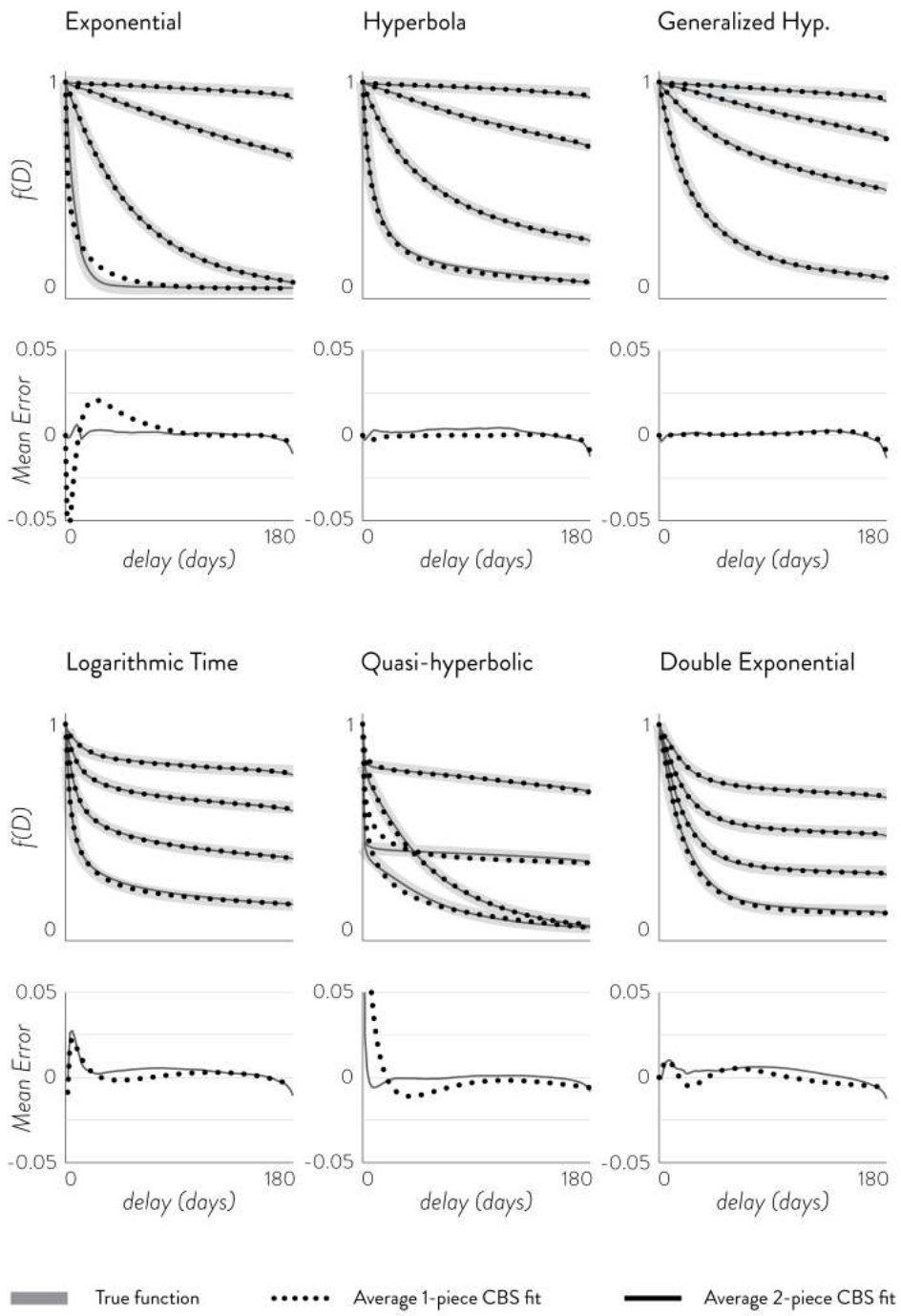
To show that CBS can nest many of the utility functions used for ITC and RC, we simulated choices from various utility functions and assessed if CBS can accurately recover them. We simulated binary choice from 6 models for ITC and 6 models for RC, each with 4 parameter combinations. The chosen models and their 4 parameter combinations are shown in **Supplemental Table 2**. Each simulated choice was between a smaller monetary amount of $20 (fixed across all trials), and a larger monetary amount that varied from trial to trial. The larger monetary amount was either delayed (for ITC models) or probabilistic (for RC models). The amount of the larger monetary option on each trial was created by uniformly sampling the ratio between the smaller and larger monetary amount (0 ~ 1; e.g., ratio of 0.5 means the smaller amount is half that of the larger amount). In ITC simulation, the delays were uniformly sampled from 0~180 days, and in RC simulation, the probabilities were uniformly sampled from 0~1. Dataset sizes ranged from $7^2$ to $20^2$ choices based on how finely we sampled the range of delay/probability and amount. The difference in utilities of the two options were used in a logit model (**eq. 3, main manuscript**) to generate choice probabilities, according to which we generated binary choices. For each of the (6+6) x 4 = 48 functions x 14 dataset size conditions ($7^2$~$20^2$), we simulated 200 datasets. The scaling parameter $\sigma$ was fixed at 1, as it is not a variable of interest in our study. After fitting each choice dataset with CBS (both 1-piece and 2-piece), 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, the MAE is best interpreted as an out-of-sample measure; 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.

CBS shows excellent recovery of various latent utility functions. **Supplemental Fig. 3** and **Supplemental Fig. 4** show the true functions and their average CBS fits from $20^2$ choice dataset conditions. In both ITC and RC, we see that both 1-piece and 2-piece CBS functions provide very close recovery of a wide variety of utility functions. Closer inspection shows that 1-piece CBS has some difficulties with sharp kinks or curves with multiple inflection points. 2-piece CBS shows near perfect approximation of all the utility functions tested. Given this result, more than 2 pieces of CBS do not seem necessary for modeling ITC or RC.

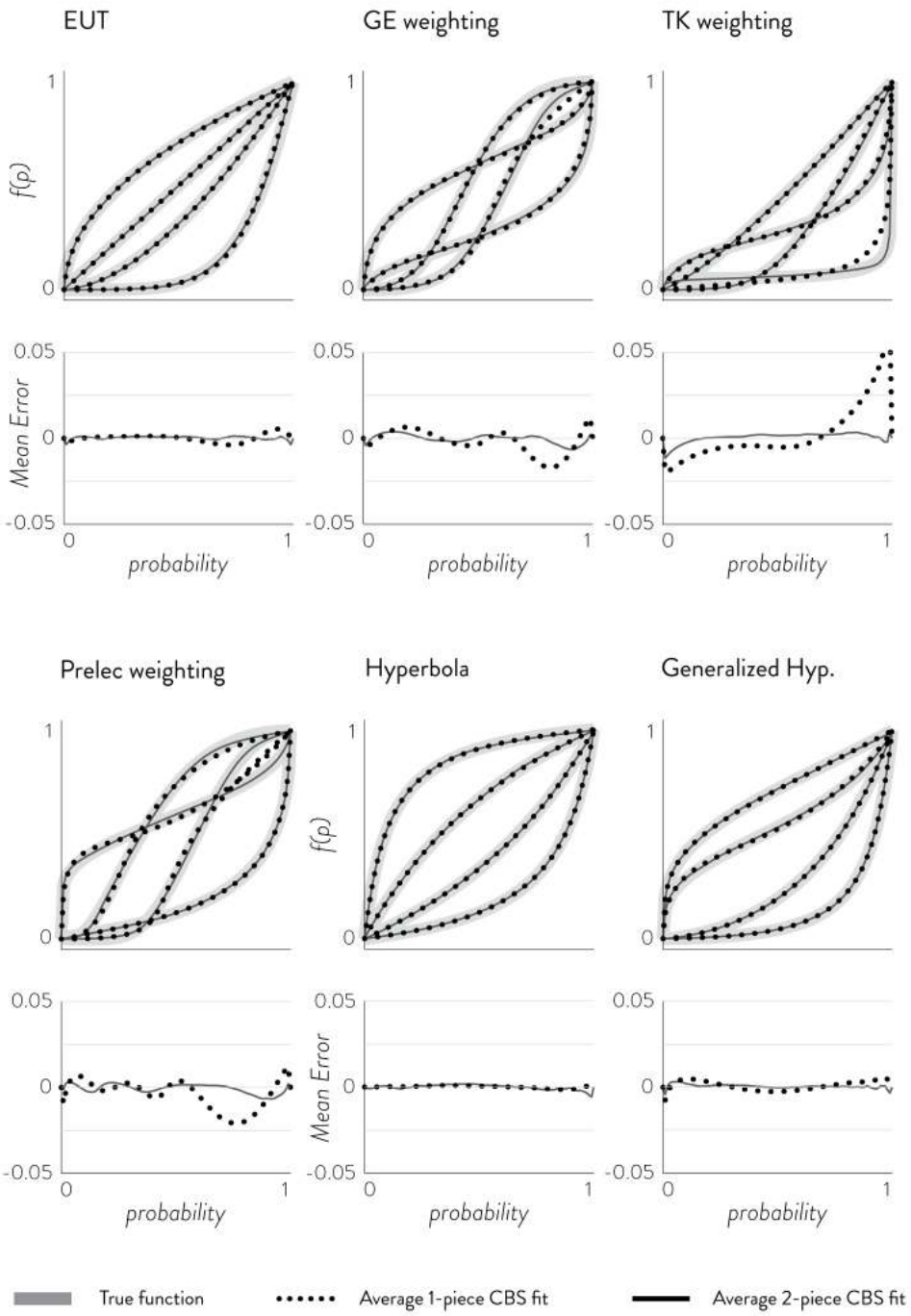
CBS shows little error even at smaller dataset sizes. **Supplemental Table 3** shows the mean MAEs of CBS fits under different simulating functions and dataset sizes. Note that this is not the MAE of the average fit (which is **Supplemental Fig. 3** and **Supplemental Fig. 4**), but rather the average MAE of the fits. The former would show CBS fits after averaging out the choice noise, but 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 0.05 for all cases, where the range of absolute errors can be between 0 and 1). Generally, we see that 1-piece CBS has lower errors than 2-piece CBS except for a few cases.

| Simulating Function | Equivalent Expression in $U = A \cdot f(X)$ form | Simulating Parameters |
|---|---|---|
| Exponential | $f(D) = \exp(-kD)$ | $\ln k \in \{-8, -6, -4, -2\}$ |
| Hyperbola | $f(D) = (1 + kD)^{-1}$ | $\ln k \in \{-8, -6, -4, -2\}$ |
| General Hyp. | $f(D) = (1 + kD)^{-s}$ | $(\ln k\,, s) \in \{(-7, 0.5), (-7, 2), (-4, 0.5), (-4, 2)\}$ |
| Logarithmic Time | $f(D) = D^{-k}$ | $k \in \{0.4, 0.2, 0.1, 0.05\}$ |
| Quasi-hyperbolic | $f(D) = \beta \exp(-kD)$ | $(\beta, \ln k) \in \{(0.4, -7), (0.4, -4), (0.8, -7), (0.8, -4)\}$ |
| Double Exp. | $f(D) = we^{-aD} + (1 - w)e^{-bD}$ | $\ln a = -8, \ln b = -3, w \in \{0.7, 0.5, 0.3, 0.1\}$ |
| EUT | $f(p) = p^{1/\alpha}$ | $\alpha \in \{0.2, 0.6, 1, 2\}$ |
| Hyperbola | $f(p) = \left(1 + h(p^{-1} - 1)\right)^{-1}$ | $h \in \{0.1, 0.5, 2, 7\}$ |
| GE weighting | $f(p) = \left(\frac{\delta p^{\gamma}}{\delta p^{\gamma} + (1 - p)^{\gamma}}\right)^{1/\alpha}$ | $\alpha = 0.8, (\delta, \gamma) \in \{(0.5,0.5), (0.5,2), (2, 0.5), (2, 2)\}$ |
| TK weighting | $f(p) = \left(\frac{p^{\gamma}}{(p^{\gamma} + (1 - p)^{\gamma})^{\frac{1}{\gamma}}}\right)^{1/\alpha}$ | $\alpha = 0.8, \gamma \in \{0.25, 0.5, 1, 3\}$ |
| Prelec weighting | $f(p) = \exp\left(-\frac{\delta}{\alpha}(-\ln p)^{\gamma}\right)$ | $\alpha = 0.8, (\delta, \gamma) \in \{(0.5,0.5), (0.5,2), (2, 0.5), (2, 2)\}$ |
| General Hyp. 1 | $f(p) = \left(1 + h(p^{-1} - 1)\right)^{-s}$ | $(h, s) \in \{(2, 0.3), (7, 0.3), (2, 1.5), (7, 1.5)\}$ |

**Supplemental Table 2. Simulating utility functions for CBS recovery.** Shown above are the ITC models (1~6) and RC models (7~12) expressed in $U = A \cdot f(X)$ form (see **Supplemental Materials Section A** for transformations). The parameter sets used to simulate choice datasets are shown on the right column.




***Supplemental Figure 3. Average CBS fits from ITC choice dataset simulation.*** *CBS fits are shown overlaid on top of 6 different ITC utility functions. 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 200 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.*




***Supplemental Figure 4. Average CBS fits from RC dataset simulation.*** *CBS fits are shown overlaid on top of 6 different RC utility functions (converted into probability space). 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 200 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.*

| Data Size | ITC Exponential | | ITC Hyperbola | | ITC Gen. Hyp. | | ITC Log Time | | ITC Quai-hyp | | ITC Double Exp. | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 49 | **.033** | .037 | **.036** | .042 | **.038** | .043 | **.040** | .047 | **.038** | .040 | **.045** | .048 |
| 64 | **.030** | .031 | **.031** | .036 | **.033** | .036 | **.033** | .044 | **.029** | .037 | **.034** | .043 |
| 81 | **.028** | .028 | **.027** | .032 | **.030** | .034 | **.026** | .035 | **.028** | .025 | **.033** | .040 |
| 100 | **.025** | .026 | **.025** | .030 | **.027** | .031 | **.025** | .032 | **.026** | .028 | **.027** | .035 |
| 121 | **.024** | .024 | **.022** | .028 | **.024** | .028 | **.022** | .032 | **.023** | .024 | **.022** | .038 |
| 144 | .023 | **.023** | **.021** | .026 | **.022** | .026 | **.021** | .031 | .022 | **.018** | **.019** | .035 |
| 169 | .021 | **.019** | **.019** | .024 | **.020** | .024 | **.018** | .031 | .021 | **.019** | **.018** | .028 |
| 196 | .020 | **.019** | **.018** | .025 | **.019** | .023 | **.017** | .035 | .020 | **.016** | **.016** | .026 |
| 255 | .019 | **.017** | **.017** | .020 | **.018** | .021 | **.016** | .027 | .019 | **.015** | **.015** | .020 |
| 256 | .019 | **.016** | **.016** | .023 | **.017** | .020 | **.014** | .023 | .018 | **.014** | **.015** | .021 |
| 289 | .018 | **.016** | **.014** | .021 | **.016** | .019 | **.014** | .019 | .018 | **.013** | **.014** | .019 |
| 324 | .018 | **.018** | **.014** | .020 | **.015** | .018 | **.013** | .019 | .017 | **.013** | **.013** | .019 |
| 361 | .017 | **.016** | **.014** | .022 | **.014** | .018 | **.012** | .019 | .017 | **.012** | **.012** | .017 |
| 400 | .017 | **.015** | **.013** | .019 | **.013** | .016 | **.012** | .017 | .016 | **.011** | **.012** | .017 |

| Data Size | RC EUT | | RC Hyperbola | | RC GE weight | | RC TK weight | | RC Prelec Weight | | RC Gen. Hyp. | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 49 | **.032** | .046 | **.038** | .051 | **.041** | .044 | **.034** | .039 | **.039** | .045 | **.037** | .046 |
| 64 | **.028** | .040 | **.032** | .042 | **.036** | .039 | **.032** | .033 | **.032** | .038 | **.030** | .040 |
| 81 | **.025** | .034 | **.027** | .036 | **.030** | .035 | **.026** | .030 | **.031** | .034 | **.025** | .033 |
| 100 | **.020** | .029 | **.023** | .030 | **.024** | .028 | **.025** | .026 | **.028** | .029 | **.021** | .030 |
| 121 | **.018** | .025 | **.020** | .027 | **.021** | .027 | .024 | **.022** | .027 | **.026** | **.019** | .026 |
| 144 | **.015** | .023 | **.019** | .026 | **.022** | .024 | .022 | **.020** | .025 | **.024** | **.016** | .024 |
| 169 | **.014** | .022 | **.017** | .024 | **.020** | .021 | .022 | **.019** | .025 | **.022** | **.014** | .022 |
| 196 | **.014** | .021 | **.016** | .022 | **.017** | .020 | .019 | **.017** | .025 | **.019** | **.014** | .021 |
| 255 | **.012** | .019 | **.015** | .020 | **.016** | .019 | .019 | **.018** | .023 | **.019** | **.013** | .019 |
| 256 | **.012** | .018 | **.014** | .019 | **.014** | .018 | .019 | **.017** | .022 | **.018** | **.013** | .017 |
| 289 | **.011** | .016 | **.013** | .018 | **.014** | .017 | .018 | **.014** | .023 | **.017** | **.012** | .016 |
| 324 | **.010** | .016 | **.012** | .017 | **.015** | .016 | .018 | **.014** | .023 | **.016** | **.011** | .015 |
| 361 | **.010** | .015 | **.011** | .016 | **.015** | .015 | .017 | **.013** | .023 | **.014** | **.010** | .015 |
| 400 | **.009** | .014 | **.011** | .016 | **.014** | .015 | .017 | **.013** | .022 | **.014** | **.010** | .014 |

**Supplemental Table 3. Mean Absolute Error (MAE) for utility function recovery with CBS.** Across various dataset sizes (left most column), the average MAE between the fitted CBS functions and the true simulating functions are shown (1-piece on left 2-piece on right). The lower MAE of the two CBS functions are bolded.

**H. Summary measure of impulsivity and risk-aversion via CBS AUC**

The area under the curve (AUC) of fitted CBS functions serves as a summary measure of impulsivity and risk-aversion. While AUC can be easily calculated numerically via approximations, the analytical expression for its calculation is quite straightforward. In the case of a 2-piece CBS function, one can simply add up the AUC of each individual piece. The formula for AUC can be calculated quite straightforwardly by applying integration on equations (1) and (2) of main manuscript. Let $(x_i, y_i)$ denote the coordinates of the point $P_i$. Then, the AUC of each CBS piece is as follows:

$$\int_{x_0}^{x_3} y\,dx = \frac{1}{20}\begin{pmatrix} 6x_1y_0 - 6x_0y_1 - 10x_0y_0 - 3x_0y_2 + 3x_2y_0 - x_0y_3 - 3x_1y_2 \\ +3_{x2}y_1 + x_3y_0 - 3x_1y_3 + 3x_3y_1 - 6x_2y_3 + 6x_3y_2 + 10x_3y_3 \end{pmatrix} \quad (37)$$

**I. Supplemental references**

Brezger, A., & Steiner, W. J. (2008). Monotonic Regression Based on Bayesian P–Splines. *Journal of Business & Economic Statistics*. https://doi.org/10.1198/073500107000000223

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

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

Leitenstorfer, F., & Tutz, G. (2007). Generalized monotonic regression based on B-splines with an application to air pollution data. *Biostatistics*. https://doi.org/10.1093/biostatistics/kxl036