



Management Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Dynamic Experiments for Estimating Preferences: An Adaptive Method of Eliciting Time and Risk Parameters

Olivier Toubia, Eric Johnson, Theodoros Evgeniou, Philippe Delquié,

To cite this article:

Olivier Toubia, Eric Johnson, Theodoros Evgeniou, Philippe Delquié, (2013) Dynamic Experiments for Estimating Preferences: An Adaptive Method of Eliciting Time and Risk Parameters. *Management Science* 59(3):613-640. <http://dx.doi.org/10.1287/mnsc.1120.1570>

Full terms and conditions of use: <http://pubsonline.informs.org/page/terms-and-conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2013, INFORMS

Please scroll down for article—it is on subsequent pages



INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Dynamic Experiments for Estimating Preferences: An Adaptive Method of Eliciting Time and Risk Parameters

Olivier Toubia, Eric Johnson

Columbia Business School, Columbia University, New York, New York 10027 {ot2107@columbia.edu, ejj3@columbia.edu}

Theodoros Evgeniou

INSEAD, 77305 Fontainebleau, France, theodoros.evgeniou@insead.edu

Philippe Delqu  

School of Business, The George Washington University, Washington, DC 20052, delquie@gwu.edu

We present a method that dynamically designs elicitation questions for estimating risk and time preference parameters. Typically these parameters are elicited by presenting decision makers with a series of static choices between alternatives, gambles, or delayed payments. The proposed method dynamically (i.e., adaptively) designs such choices to optimize the information provided by each choice, while leveraging the distribution of the parameters across decision makers (heterogeneity) and capturing response error. We explore the convergence and the validity of our approach using simulations. The simulations suggest that the proposed method recovers true parameter values well under various circumstances. We then use an online experiment to compare our approach to a standard one used in the literature that requires comparable task completion time. We assess predictive accuracy in an out-of-sample task and completion time for both methods. For risk preferences, our results indicate that the proposed method predicts subjects' willingness to pay for a set of out-of-sample gambles significantly more accurately, while taking respondents about the same time to complete. For time preferences, both methods predict out-of-sample preferences equally well, while the proposed method takes significantly less completion time. For risk and time preferences, average completion time for our approach is approximately three minutes. Finally, we briefly review three applications that used the proposed methodology with various populations, and we discuss the potential benefits of the proposed methodology for research and practice.

Key words: prospect theory; time discounting; Bayesian statistics; adaptive experimental design; revealed preference

History: Received June 28, 2011; accepted April 1, 2012, by Teck Ho, decision analysis. Published online in *Articles in Advance* October 8, 2012.

1. Introduction

Often, in trying to understand choices, we would like to know the decision maker's underlying preferences. The development of behavioral-based models of preferences, such as prospect theory (Kahneman and Tversky 1979, Tversky and Kahneman 1992) or time discounting models (Frederick et al. 2002 and references therein) has inspired an enormous amount of experimental and observational research in both laboratory and field settings (e.g., Barberis et al. 2001, Camerer et al. 2003, Harrison et al. 2002, Tanaka et al. 2010, and references therein). These papers typically elicit the parameters of a model (e.g., prospect theory), by asking subjects to make decisions about alternatives (e.g., gambles), and then use these parameters to reach empirical conclusions about people's behavior. For example, the recent work of Tanaka et al. (2010) examines how Vietnamese villagers' risk and

time preferences relate to assorted socioeconomic variables and choices. The feasibility of such experimental or field studies and their empirical conclusions depends on both the quality of the parameter estimates and the time required to complete the elicitation tasks. As these studies become larger, use more diverse sets of participants and move into settings where there is less control (e.g., online studies or field studies in the developing world), improving the accuracy and time efficiency of the elicitation methods become paramount.

In this paper, we present a novel parametric dynamic experiments for estimating preferences (DEEP) methodology. Our main focus is the elicitation of the parameters of cumulative prospect theory (CPT) (Tversky and Kahneman 1992) and a quasi-hyperbolic time discounting (QTD) model (Benhabib et al. 2010, Frederick et al. 2002, Laibson 1997, Phelps

and Pollak 1968). However, our methodology is quite general and can be adapted to other functional forms. To demonstrate, we also apply our method to the estimation of a sign-dependent CPT model, the time discounting model proposed by Ebert and Prelec (2007), and a model recently proposed by Baucells and Heukamp (2012) that jointly captures risk and time preferences. Our work relies on concepts from the preference measurement literature coming predominantly from marketing (e.g., Allenby and Rossi 1999, Lenk et al. 1996, Green and Rao 1972, Rossi and Allenby 2003, Srinivasan and Shocker 1973), and bridges that literature and the preference assessment literature in decision theory.

Our main contribution is methodological. We elicit risk and time preferences by dynamically (i.e., adaptively) optimizing the sequence of questions presented to each subject, while leveraging information about the distribution of the parameters across individuals (heterogeneity) and modeling response error explicitly. Once the data have been collected, we recommend combining our adaptive questionnaire design method with hierarchical Bayes, a well-proven method for estimating the parameters given the data (e.g., Jarnebrant et al. 2009, Nilsson et al. 2011). Our method is computationally efficient and easy for the experimenter because it automatically precalculates a table of all possible question paths, which can be used as a lookup table during the study, as done in our online experiment.¹

We study the convergence of our approach using a series of simulations. Our simulations cover five preference models (i.e., CPT, sign-dependent CPT, QTD, the Ebert–Prelec model, and the Baucells–Heukamp model). We then compare our methodology to a titration method routinely used in psychology and marketing (Weber et al. 2007, Zauberman 2003) and experimental economics, where it is often called a price-list method (Ashraf et al. 2006, Harrison et al. 2002, Meier and Sprenger 2009), using an online experiment. The specific titration method we use as a benchmark is adapted from Tanaka et al. (2010), because it is one of the few studies that assess models of both risk and time, and has comparable completion times. Following Tanaka et al. (2010), our online experiment focuses on estimating the parameters of the CPT and QTD models for more than 500 individuals. For risk preferences, our method performs significantly better on out-of-sample predictions and requires similar response time; for time preferences, our method performs similarly well on out-of-sample predictions and significantly better on response time.

Each elicitation task requires, on average, less than three minutes. The proposed methodology produces parameter values that are in the range reported in the literature. We also briefly review other studies that have used the proposed methodology with various populations. These studies further support our empirical findings of fast completion times and estimates being in the range of those reported in the literature, and illustrate how the proposed method may enable researchers to uncover relations between time and risk preferences and other covariates or behaviors.

The paper is organized as follows. After reviewing extant-related methodologies and introducing the notation, we present our general framework in §2 and our specific questionnaire design method in §3. We explore the convergence of our approach using simulations in §4. We compare our approach to a benchmark adapted from Tanaka et al. (2010) using an online experiment in §5. We conclude in §6, where we also discuss some benefits of the proposed method and briefly review three other recent studies in which it has been used.

1.1. Related Methodologies

A number of methodologies for measuring parameters of preference models such as CPT or QTD have been developed in decision analysis (e.g., Abdellaoui et al. 2008, Wakker and Deneffe 1996, among others). These methods differ in many respects, including the type of elicitation responses (e.g., payoff or probability), whether they use choices or indifference judgments, whether the questions are chained, whether parametric forms are assumed, and so on. However, few of the parametric methods explicitly leverage the distribution of the parameters across individuals (heterogeneity) to inform and improve individual-level parameter estimates. Exceptions include the recent work by Bruhin et al. (2010), who apply latent-class analysis to the estimation of probability distortion,² and Jarnebrant et al. (2009) and Nilsson et al. (2011), who estimate prospect theory parameters using hierarchical Bayes methods. By contrast, our approach captures and leverages heterogeneity in parameter values both when designing questionnaires and when estimating the parameters. Moreover, traditional methods often use adaptive questions, called staircase methods in psychophysics, and typically employ a bisection search process to zero in on a point or small interval where preference switches. The questions in our designs are adaptively developed with the only restriction from previous questions being that each new question adds the most information

¹ This table and the code that was used to create it, estimate the parameters, and construct the online interface are publicly available from the authors upon request.

² Latent class analysis has a long history in marketing (see Kamakura and Russell 1989 for one of the earlier applications). Comparisons of latent class with hierarchical Bayes approaches as we use here, have suggested that both fit the data equally well overall (see, e.g., Andrews et al. 2002a, b).

in a certain statistical sense, given the respondent's responses to the previous questions.

One recent approach that is comparable to ours is that of Wang et al. (2010). These authors use a similar principle of selecting questions adaptively to maximize information, where information is measured using the Kullback–Leibler divergence. One key difference between that and our approach is that the questionnaire design approach proposed by Wang et al. (2010) requires discretizing the distribution of the parameters. For example, in one of their studies, the authors design questions assuming that each parameter may take 10 possible values. In contrast, our questionnaire design approach is designed to deal with continuous parameter spaces. This allows dealing with continuous distributions on the parameters and infinite sets of possible parameter values, and to consider models with more parameters (in the Wang et al. (2010) approach, the number of possible sets of values grows exponentially in the number of parameters). Wang et al. (2010) apply their approach to the simultaneous estimation of risk aversion and loss aversion (two parameters). In contrast, we apply our approach to the estimation of five different risk and time preference models, where the number of parameters estimated per decision maker goes up to five (and could be more for other models). Another recent paper related to ours is the paper by Cavagnaro et al. (2012), who develop an adaptive approach for model discrimination (e.g., determining whether a subject's choices are more compatible with one model or another). In contrast, our approach estimates preference parameters given an assumed model of preferences. The model selection problem is discrete in nature (there is a discrete set of candidate models), whereas ours is continuous (given a model, the set of possible parameter values is continuous), giving rise to different approaches.

The dynamic questionnaire design approach used in this paper is inspired by adaptive design methods recently developed in the preference measurement literature (Abernethy et al. 2008; Sawtooth Software 1996; Toubia et al. 2003, 2004, 2007a, b). These methods, which optimize experimental designs dynamically, have been used widely and successfully for preference measurement in marketing both by researchers and practitioners. However, they assume that utility is linear in the preference parameters, and are therefore not directly applicable to the measurement of risk and time preferences, which typically relies on highly nonlinear functional forms. The adaptive preference measurement framework we develop here is based on a Bayesian framework and accommodates any functional form, including highly nonlinear ones, as long as they are twice differentiable in the preference parameters. Therefore, in addition

to the five different risk and time preference models considered here, the framework can be applied to other preference models as well.

1.2. Background and Notation

For simplicity, we discuss the methodology in §§2 and 3 using a general notation that applies to any parametrically specified model of risk and/or time preferences. This creates a compact notation with some overlap for the various models that we clarify as needed.

1.2.1. Risk Preferences and Cumulative Prospect Theory. Prospect theory (Kahneman and Tversky 1979) and its extension CPT (Tversky and Kahneman 1992) are widely used descriptive models of choice under risk. CPT has three main features: a value function defined on gains and losses, which accounts for the fact that people are sensitive to changes in wealth rather than total wealth; loss aversion, which reflects that people are more sensitive to losses than to gains of the same magnitude; and probability weighting, which captures the fact that people tend to weigh probabilities in a nonlinear fashion, particularly near certainty. We apply our approach to a version of CPT, where the probability weighting and the curvature of the value function are identical for gains and losses (as in other work, e.g., Tanaka et al. 2010) and to a version where they both differ (Tversky and Kahneman 1992).

The gambles we use are defined by $\{x, p; y\}$ such that the outcome of the gamble is x with probability p , and y with probability $1 - p$. We assume that the CPT probability weighting function is as proposed by Prelec (1998). Therefore, a decision maker's preferences for gambles are defined by three parameters $\{\alpha, \sigma, \lambda\}$ in the simpler version of the model, and by five parameters $\{\alpha^+, \alpha^-, \sigma^+, \sigma^-, \lambda\}$ for the sign-dependent version of the model. These parameters capture, respectively, the distortion of probabilities, the curvature of the value function, and loss aversion. Formally, in CPT, the value of a gamble to an individual is given by (without loss of generality, we assume $|x| > |y|$ when x and y have the same sign; otherwise, x and y may be swapped):

CPT:

$$U(x, p, y, \alpha, \sigma, \lambda) = \begin{cases} v(y, \sigma) + \pi(p, \alpha)(v(x, \sigma) - v(y, \sigma)) & \text{if } x > y > 0 \text{ or } x < y < 0, \\ \pi(p, \alpha)v(x, \sigma) + \pi(1 - p, \alpha)v(y, \sigma) & \text{if } x < 0 < y, \end{cases}$$

where

$$v(x, \sigma) = \begin{cases} x^\sigma & \text{for } x > 0, \\ -\lambda(-x)^\sigma & \text{for } x < 0, \end{cases}$$

and $\pi(p, \alpha) = \exp[-(-\ln p)^\alpha]$.

The sign-dependent version of the model is as follows:

Sign-dependent CPT:

$$U(x, p, y, \alpha^+, \alpha^-, \sigma^+, \sigma^-, \lambda) = \begin{cases} v(y, \sigma^+, \sigma^-) + \pi(p, \alpha^+) \cdot (v(x, \sigma^+, \sigma^-) - v(y, \sigma^+, \sigma^-)) & \text{if } x > y > 0, \\ v(y, \sigma^+, \sigma^-) + \pi(p, \alpha^-) \cdot (v(x, \sigma^+, \sigma^-) - v(y, \sigma^+, \sigma^-)) & \text{if } x < y < 0, \\ \pi(p, \alpha^-)v(x, \sigma^+, \sigma^-) + \pi(1 - p, \alpha^+)v(y, \sigma^+, \sigma^-) & \text{if } x < 0 < y, \end{cases}$$

where

$$v(x, \sigma^+, \sigma^-) = \begin{cases} x^{\sigma^+} & \text{for } x > 0, \\ -\lambda(-x)^{\sigma^-} & \text{for } x < 0, \end{cases}$$

and $\pi(p, \alpha) = \exp[-(-\ln p)^\alpha]$.

We elicit the CPT parameters by asking decision makers to make a series of choices between pairs of gambles. We index decision makers by i ($i = 1, \dots, I$) and denote by w_i the vector of parameters for decision maker i : $w_i = [\alpha_i; \sigma_i; \lambda_i]$ for CPT, and $w_i = [\alpha_i^+; \alpha_i^-; \sigma_i^+; \sigma_i^-; \lambda_i]$ for sign-dependent CPT. We index questions by j ($j = 1, \dots, J$), such that question j for respondent i consists in choosing between gamble $X_{ij}^A = \{x_{ij}^A, p_{ij}^A; y_{ij}^A\}$ and gamble $X_{ij}^B = \{x_{ij}^B, p_{ij}^B; y_{ij}^B\}$.

1.2.2. Time Preferences. In experimental studies of time preferences, subjects are typically faced with choices between a smaller-sooner reward and a larger-later reward. The choice alternatives take the form $\{x, t\}$, meaning a reward x to be received t periods (e.g., days) from now. The models we consider to represent preferences for payoffs occurring in time are discounted utility models $U(x, t) = v(x)d(t)$, where v is the utility of receiving reward x and d is the discount function—as noted above, other models can be used. By and large, the literature on delayed reward preferences is concerned with the shape and nature of the discount function d . Classic forms are exponential discounting (the standard model in economics, with constant discount rate) and hyperbolic discounting, which implies a discount rate decreasing with time. Although our approach may be extended to other functional forms, here, we consider two time preference models: a “quasi-hyperbolic” discount function (Angeletos et al. 2001, Benhabib et al. 2010, Frederick et al. 2002, Laibson 1997, Phelps and Pollak 1968) and the model proposed by Ebert and Prelec (2007). In both cases, we assume a linear value function of payoff.³

Specifically, the QTD model we estimate is of the following form (Benhabib et al. 2010, Laibson 1997, Phelps and Pollak 1968):

$$\text{QTD: } U(x, t, \beta, r) = xd(t, \beta, r),$$

where

$$d(t, \beta, r) = \begin{cases} 1 & \text{for } t = 0, \\ \beta \exp(-rt) & \text{for } t > 0. \end{cases}$$

For $\beta < 1$, the discount function presents a discontinuous drop at $t = 0$, which reflects the empirical observation that the present $t = 0$ is overweighted relative to any future $t > 0$. This is also called *present bias* (O'Donoghue and Rabin 1999).

The Ebert–Prelec model is specified as follows:

$$\text{Ebert–Prelec: } U(x, t, b, r) = xd(t, b, r),$$

where

$$d(t, b, r) = \exp(-(rt)^b).$$

When $b = 1$, the Ebert–Prelec model is similar to the exponential discounting model. One interesting feature of this model is its resemblance with the Prelec probability weighing function. Baucells and Heukamp (2012) proposed a model that links probability weighing and time discounting, as reviewed next.

We elicit the parameters of the QTD model $w_i = [\beta_i; r_i]$ or of the Ebert–Prelec model $w_i = [b_i; r_i]$ of decision maker i through a series of choices between pairs of delayed payments—where the delay of an immediate payment is zero. Question j for respondent i consists of choosing between $X_{ij}^A = \{x_{ij}^A, t_{ij}^A\}$ and $X_{ij}^B = \{x_{ij}^B, t_{ij}^B\}$.

1.2.3. Joint Estimation of Risk and Time Preferences. Finally, we show how our approach may be applied to a model that jointly captures risk and time preferences. We illustrate this with the “power” model proposed by Baucells and Heukamp (2012). Here, the choice alternatives take the form $\{x, p, t\}$, meaning a reward x to be received t periods (e.g., days) from now with probability p , and no reward received with probability $1 - p$. The model we consider here represents preferences for such gambles as

experiment, to make the comparison with the method of Tanaka et al. (2010) cleaner, we estimated CPT and QTD preferences separately and assumed a linear value function when estimating time preferences.

³ This linearity assumption is not necessary. We consider next a model that jointly captures time and risk preferences. In our

$U(x, p, t) = v(x)d(p, t)$. Precisely, the model is specified as follows:⁴

Baucells–Heukamp:

$$U(x, p, t, b, r, \sigma) = x^\sigma d(p, t, b, r),$$

where $d(p, t, b, r) = \exp(-(-\ln p + rt)^b)$.

This model nests the Ebert–Prelec model for sure gambles ($p = 1$) if $\sigma = 1$ and nests the Prelec probability weighing function when $t = 0$. Baucells and Heukamp (2012) interpret the parameter b as capturing sensitivity to psychological distance, where psychological distance is larger for outcomes that are less likely and/or more distant in time (Trope et al. 2007). They interpret the parameter r as a *probability discount rate*, which measures the trade-off between probability and time delay. We elicit the parameters of the Baucells–Heukamp model $w_i = [b_i; r_i; \sigma_i]$ of decision maker i through a series of choices between pairs of delayed gambles. Question j for respondent i consists of choosing between $X_{ij}^A = \{x_{ij}^A, p_{ij}^A, t_{ij}^A\}$ and $X_{ij}^B = \{x_{ij}^B, p_{ij}^B, t_{ij}^B\}$.

2. Framework

Our adaptive questionnaire design method is based on a Bayesian framework. The two main building blocks of this framework are the treatment of response errors, and the specification of a prior distribution of the parameters. We introduce each of these two features in sequence. Note that although we assume pairwise choice data here, the methodology can be applied for other types of questions (e.g., choices between more than two alternatives, certainty equivalents, willingness to pay, etc.). Note also that our Bayesian framework nests a frequentist framework. In particular, if a uniform prior distribution is assumed in our Bayesian framework, the posterior distribution of the parameters becomes the likelihood function that would be used in a frequentist framework. Using a nonuniform prior not only enables us to leverage the distribution of the parameters across decision makers, it also increases the robustness of the questionnaires to response error. Indeed, the use of a prior reduces overfitting by shrinking the estimates toward the mean of the prior, which is particularly relevant in the presence of response error and/or when the number of questions asked is small (Evgeniou et al. 2007).

2.1. Response Error

We assume that faced with a choice between two options (gambles or delayed payments), a decision

maker will not systematically choose the one with the higher value. Such deviations may be interpreted as being the result of unobservable perturbations to the decision makers' preferences, or simply response error. The existence of noise in the decision makers' choices has long been recognized in the literature (e.g., Luce 1958, Laskey and Fischer 1987). There are many ways to model response error (e.g., Hey and Orme 1994). We make a particular choice here, but other models could be used. We introduce response error by modeling the probability that decision maker i chooses option A over option B in question j using a logistic specification common in choice modeling, and used previously in the estimation of risk and time preference parameters (see, e.g., Tanaka et al. 2010, Tom et al. 2007):

$$P_{ij} = \frac{\exp(\delta U(X_{ij}^A, w_i))}{\exp(\delta U(X_{ij}^A, w_i)) + \exp(\delta U(X_{ij}^B, w_i))}. \quad (1)$$

The parameter δ captures the amount of response error (equivalent to a logit scale parameter). Higher values of δ imply less response error (the choice probabilities converge to 0 or 1). For ease of exposition, after each question, we relabel the chosen option as option 1 and the other as option 2.

2.2. Prior Distributions

We formulate a Bayesian prior distribution on w_i . Such a prior allows capturing any prior information about these parameters, and allows estimating the parameters of each individual separately while leveraging the distribution of parameters across individuals. Although any prior distribution may be used, the most common in the preference measurement literature is the normal distribution (often truncated): $w_i \sim \text{TN}(w_0, D)$, where we truncate the normal distribution to ensure that the parameters remain in an acceptable range (for CPT and sign-dependent CPT, we impose $\alpha_i \in [0.05, 2]$, $\sigma_i \in [0.05, 2]$, and $\lambda_i \in [0, 10]$; for QTD, we impose $\beta_i \in [0, 2]$ and $r_i \in [0, 0.05]$;⁵ for the Ebert–Prelec model, we impose $b_i \in [0.05, 2]$ and $r_i \in [0, 0.05]$; for the Baucells–Heukamp model, we impose $b_i \in [0.05, 2]$, $r_i \in [0, 0.05]$, and $\sigma_i \in [0.05, 2]$).

Intuitively, the prior distribution effectively shrinks w_i toward a common vector w_0 (the “average” from which everyone deviates). The amount of shrinkage is governed by the covariance matrix of the prior distribution, D .⁶ Using Bayes theorem, the prior distribution on w_i is combined with the likelihood in Equation (1) to obtain a posterior distribution on w_i .

⁵ Time is measured in days in our implementation.

⁶ The matrix D is a $n_{\text{par}} \times n_{\text{par}}$ matrix and the vector w_0 is a n_{par} -dimensional vector, where n_{par} is the number of parameters in the model. For QTD and Ebert–Prelec, $n_{\text{par}} = 2$; for CPT and Baucells–Heukamp, $n_{\text{par}} = 3$; for sign-dependent CPT, $n_{\text{par}} = 5$.

⁴ Baucells and Heukamp (2012) allow for the parameter r to depend on x , as a way to capture magnitude effects in time preferences. For simplicity, we set r constant.

In this paper, we consider two types of priors. The simplest type is one that is fixed before collecting the data. We use this type of prior in our adaptive questionnaire design, for computational reasons. Another option is to formulate prior distributions on the priors themselves. This approach, called *hierarchical Bayes*, allows the priors to be influenced by the data. We recommend using hierarchical Bayes to estimate the parameters once the data have been collected, as we do in our online study. In our simulations where we estimate parameters for one subject at a time, we use fixed priors for questionnaire design as well as estimation.

2.2.1. Fixed Priors (Used for Questionnaire Design). When fixed prior values are used for w_0 , D , and δ , denoted as \hat{w}_0 , \hat{D} , and $\hat{\delta}$, the posterior distribution of the parameters for decision maker i is simply given by Bayes' rule as follows:

$$P(w_i | data, \hat{w}_0, \hat{D}, \hat{\delta}) \propto P(data | w_i, \hat{\delta})P(w_i | \hat{w}_0, \hat{D}),$$

where in our case, $P(data | w_i, \hat{\delta})$ is given by Equation (1) and $P(w_i | \hat{w}_0, \hat{D}) = \text{TN}(\hat{w}_0, \hat{D})$. If respondent i has answered q questions, this gives rise to (recall that we relabel the chosen option in each question as option 1)

$$\begin{aligned} P(w_i | data, \hat{w}_0, \hat{D}, \hat{\delta}) \\ \propto \prod_{j=1}^q \frac{\exp(\hat{\delta}U(X_{ij}^1, w_i))}{\exp(\hat{\delta}U(X_{ij}^1, w_i)) + \exp(\hat{\delta}U(X_{ij}^2, w_i))} \\ \cdot \exp\left(-\frac{1}{2}(w_i - \hat{w}_0)^T \hat{D}^{-1}(w_i - \hat{w}_0)\right). \end{aligned} \quad (2)$$

The values of \hat{w}_0 , \hat{D} , and $\hat{\delta}$ could be obtained from a pretest with a relatively small number of respondents (Toubia et al. 2007b), from prior beliefs (Huber and Zwerina 1996, Sándor and Wedel 2001), or from previous studies (e.g., Wu and Gonzalez 1996 for CPT parameters). One option is also to select an uninformative prior, by setting \hat{D} to a very “large” matrix. In this paper, we used values based on the literature for \hat{w}_0 in CPT and QTD: $\hat{w}_0 = [0.6; 0.8; 2.2]$ and $[0.8; 0.008]$ for CPT and QTD, respectively. In our simulations, we set \hat{w}_0 for the other models based on the CPT and QTD values.⁷ When designing questions in our online experiment, we assumed uninformative priors by using large diagonal matrices for \hat{D} (we used

$\hat{D} = 100I$, where I is the identity matrix) to ensure that the comparisons with the benchmark were not driven by the fact that our method allows for informative priors and the benchmark does not. This also makes the results of our comparisons more conservative. In our simulations (which were run after our online study), we set \hat{D} to a diagonal matrix based on our online study, leveraging the similarities across models.⁸ Finally, in our online experiment, we use $\hat{\delta} = 1$ for CPT and $\hat{\delta} = 0.05$ for QTD.

It is important to note that the fact that the prior distribution on w_i is normal does not imply that the final estimates will follow a normal distribution. Because the prior is combined with the likelihood, the shape of the posterior distribution does not necessarily coincide with the shape of the prior distribution. Note also that changing the prior to a uniform distribution would make the posterior distribution in Equation (2) equivalent to a frequentist likelihood function without any shrinkage toward a prior. Finally, other priors, such as a lognormal one, can also be used.

2.2.2. Hierarchical Priors and Hierarchical Bayes Estimation (Recommended for Estimation). Hierarchical Bayes allows capturing uncertainty on w_0 and D by treating them as random variables themselves. A prior distribution is specified on w_0 and D by the researcher, and a posterior distribution is obtained for these parameters by combining this prior with the data (using Bayes theorem). In other words, a prior distribution on the parameters of the prior distribution themselves is formulated, hence the hierarchical nature of the model. The prior on w_i is referred to as the first-stage prior and the priors on w_0 and D are the second-stage priors. The second-stage priors are usually selected to be as uninformative as possible to let the value of w_0 and D be determined primarily by the data (see below).

Hierarchical Bayes is particularly useful with panel data in which the number of units in the panel (e.g., number of respondents) is relatively large, and the data from each unit (e.g., choices between gambles, delayed payments, or delayed gambles) tend to be few and/or noisy. Given a recent trend toward larger sample sizes in decision analysis (e.g., Jarnebrant et al. 2009; Liu and Huang 2011; Meier and Sprenger 2009, 2010; Tanaka et al. 2010), we believe that hierarchical Bayes has become increasingly relevant to the estimation of risk and time preferences. Therefore,

⁷ We set \hat{w}_0 to $[0.7; 0.5; 0.7; 0.9; 2.2]$, $[0.6; 0.008]$, and $[0.6; 0.008; 0.8]$ for sign-dependent CPT, Ebert-Prelec, and Baucells-Heukamp, respectively. For sign-dependent CPT, we introduce arbitrary differences between α^+ and α^- , and between σ^+ and σ^- . We base the priors for the Ebert-Prelec and Baucells-Heukamp models on the priors for the CPT and QTD models. In particular, parallels may be made between the probability weighing function parameter α in CPT (respectively, the discount rate r in QTD) and the parameter b (respectively, parameter r) in the Ebert-Prelec and Baucells-Heukamp models.

⁸ We use $\hat{D} = \text{diag}([0.05; 0.05; 1])$, $\text{diag}([0.05; 0.05; 0.05; 0.05; 1])$, $\text{diag}([0.1; 0.0001])$, $\text{diag}([0.05; 0.0001])$, and $\text{diag}([0.05; 0.0001; 0.05])$ for CPT, sign-dependent CPT, QTD, Ebert-Prelec, and Baucells-Heukamp, respectively ($\text{diag}(x)$ is a diagonal matrix based on vector x). These values were loosely based on the diagonal elements of the covariance matrices estimated in our online study for CPT and QTD.

although we use fixed priors when designing questionnaires, we recommend estimating the parameters using hierarchical Bayes once the data have been collected. Interested readers are referred to Jarnebrant et al. (2009) and Nilsson et al. (2011) for previous applications of hierarchical Bayes to the estimation of risk preferences. This estimation method may be used with other questionnaire design methods as well. We assume that we have responses to J choice questions from I respondents as noted above. Instead of estimating the parameters corresponding to each decision maker independently as in traditional decision-analysis methods, hierarchical Bayes estimates the parameters for all individuals simultaneously. Such simultaneous estimation has been shown to lead to considerable improvements in estimation accuracy in other preference measurement setups (Allenby and Rossi 1999, Rossi and Allenby 2003, Rossi et al. 2005).

The likelihood function (capturing the link between value and response probabilities), the first-stage prior (capturing similarities across decision makers), and the second-stage prior (prior distribution on the parameters of the first-stage prior) are as follows:

Likelihood:

$$\prod_{i,j} \frac{\exp(\delta U(X_{ij}^1, w_i))}{\exp(\delta U(X_{ij}^1, w_i)) + \exp(\delta U(X_{ij}^2, w_i))}$$

First-stage prior: $w_i \sim \text{TN}(w_0, D)$
 δ : diffuse (improper) on \mathbb{R}^+

Second-stage prior: w_0 : diffuse (improper) on $\mathbb{R}^{+n_{\text{par}}}$
 $D \sim \text{IW}(\eta_0, \eta_0 \Delta_0)$,

where n_{par} is the number of parameters in the model ($n_{\text{par}} = 2$ for QTD and Ebert–Prelec, $n_{\text{par}} = 3$ for CPT and Baucells–Heukamp, $n_{\text{par}} = 5$ for sign-dependent CPT) and IW is the inverse Wishart distribution. With the exception of the specification of the likelihood function, which is specific to risk and time preferences, the specifications of all our distributions are standard in the hierarchical Bayes literature (see, e.g., Rossi and Allenby 2003, Rossi et al. 2005). We select an inverse Wishart distribution on D because it is conjugate with the likelihood function implied by $w_i \sim \text{TN}(w_0, D)$, i.e., the posterior distribution on D is inverse Wishart as well. We use diffuse improper priors on δ and w_0 (i.e., the priors on these parameters are completely “flat” and do not favor any specific value; see Appendix A for details).

Hierarchical Bayes estimation consists of sampling from the posterior distribution of the parameters. The posterior distribution is simply given by Bayes’ rule:

$$P(\{w_i\}, w_0, D, \delta \mid \text{data}) \propto P(\text{data} \mid \{w_i\}, \delta) \cdot P(\{w_i\} \mid w_0, D) \cdot P(w_0) \cdot P(D) \cdot P(\delta),$$

where $P(\text{data} \mid \{w_i\}, \delta)$ is given by the likelihood function; $P(\{w_i\} \mid w_0, D)$ is given by the first-stage

prior; and $P(w_0)$, $P(D)$, and $P(\delta)$ are the priors on w_0 , D , and δ , respectively. Drawing from this posterior distribution is achieved by using a Markov Chain Monte Carlo (MCMC) algorithm. Details are provided in Appendix A. The MCMC provides a set of values drawn from the posterior distribution, which may be used to produce point estimates of the parameters, or to make other types of inference. Point estimates, on which our analyses below are based, are typically obtained by averaging the draws from the MCMC, which approximates the mean of the posterior distribution.

As a more general setup, it is also possible that the similarities across decision makers be driven by similarities in covariates that influence the preference model parameters. For example, Tanaka et al. (2010) explore the relation between the CPT parameters and various demographic variables for Vietnamese villagers. In situations in which a given set of covariates are thought to influence w_i , the prior distribution on w_i may be replaced with (see, e.g., Allenby and Ginter 1995, Lenk et al. 1996) $w_i \sim \text{TN}(\Theta z_i, D)$, where z_i is a set of covariates for respondent i , and Θ is a matrix capturing the relationships between these covariates and the mean of the first-stage prior. This matrix may be fixed (if fixed priors are used) or estimated (if hierarchical priors are used). Details are provided in Appendix A. For ease of exposition, we focus here on the case in which covariates are *not* used.

3. Dynamic Questionnaire Design

Our approach is based on principles from the experimental design literature (Ford et al. 1989, Kuhfeld et al. 1994, McFadden 1974, Steinberg and Hunter 1984, and references therein). These principles have been used to develop dynamic methodologies for preference elicitation before, for example, for conjoint analysis in Marketing (Abernethy et al. 2008; Sawtooth Software 1996; Toubia et al. 2003, 2004, 2007a, b). We also rely on the Bayesian experimental design literature (see, e.g., Chaloner and Verdinelli 1995). Before describing the method in details, we present and explain the main criterion that constitutes the foundations of the method.

3.1. Questionnaire Design Criterion

Let us consider a decision maker who has responded to q binary choice questions. Our challenge is to construct the $(q + 1)$ th question for that decision maker, which will consist of two items: $X_{i(q+1)}^A$ and $X_{i(q+1)}^B$. Items A and B will be labeled 1 and 2 after the decision maker’s choice, i.e., the preferred alternative will be relabeled as 1. The central idea of our method, used in the rich literature on experimental design, is to design questionnaires such that the asymptotic covariance matrix of the maximum

likelihood estimate (MLE) of the relevant parameters is as “small” as possible, according to some defined measure. Intuitively, this ensures that the parameters are elicited with as little uncertainty as possible. For example, in the unidimensional case, the covariance matrix is simply the variance of that estimate, which governs the confidence interval around the estimate. In the multidimensional case, the covariance matrix governs the size and shape of the confidence ellipsoid around the parameter estimates (Greene 2000).

It has been shown (see, e.g., McFadden 1974, Newey and McFadden 1994) that under general conditions, the asymptotic covariance matrix of the MLE is equal to the inverse of the Hessian (i.e., second derivative matrix) of the log-likelihood function (taken at the MLE). Therefore, reducing the asymptotic covariance matrix of the MLE is achieved by maximizing some norm of the Hessian of the likelihood function. Intuitively, doing so decreases our uncertainty on the decision maker’s parameters. Different norms can and have been used, such as the absolute value of the determinant (the norm we use in this paper), the absolute value of the largest eigenvalue, the trace norm, etc.

Given our Bayesian framework, a design criterion using the same insight is that each new question should maximize the norm of the Hessian of the posterior distribution at its mode. The mode of the posterior distribution, also called the *maximum a posteriori estimate* (De Groot 1970), becomes the standard MLE in the case of a uniform prior (or if $\hat{D} \rightarrow \infty$).

3.2. Technical Details of the Method

Implementing the design criterion outlined above requires performing the following computations between the q th and $(q+1)$ th question for each decision maker and for each value of q : (i) identify the mode of the posterior distribution (the posterior distribution changes after each new question), (ii) identify the question that maximizes the expected value after the $(q+1)$ th question of a norm of the Hessian of the posterior at its mode. These computations are efficiently performed as follows:

(i) *Identify the mode of the posterior distribution.* Given the assumed prior distribution, the posterior likelihood on decision maker i ’s parameters after q questions is given by Equation (2). The mode of this posterior may be computed very quickly by maximizing the log of this expression using Newton’s method. Let \hat{w}_{iq} be the mode of the posterior based on q questions.

(ii) *Identify the question that maximizes a norm of the Hessian.* We refer to the set of possible questions as “candidate” questions. These questions consist of all possible pairs of alternatives (gambles, delayed payments, or delayed gambles) from a candidate set. This set of inputs depends on the

domain and range of payoffs over which preferences are to be elicited. This consideration is a premise to any preference assessment method. For CPT and sign-dependent CPT, we use fractional factorial sets of gambles as candidates, with a range of outcomes similar to the one used by Tanaka et al. (2010) (with a thousand Vietnamese dong replaced with one U.S. dollar). In particular, we used subsets of all gambles $\{x, p, y\}$, where $x \in \{1, 30, 40, 100, 1,000\}$, $p \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, and $y \in \{-20, -15, -10, -5, 5, 10, 30\}$, with x and y in U.S. dollars. For QTD and Ebert–Prelec, we used a fractional factorial set of delayed payments (x, t) , where $x \in \{5, 10, 15, 20, 30, 50, 80, 95, 96, 97, 98, 99, 100, 120, 150, 245, 246, 247, 248, 249, 250, 300\}$ and $t \in \{0, 3, 7, 14, 30, 60, 90\}$, with x in U.S. dollars and t in days.⁹ For Baucells–Heukamp, we used a fractional factorial set of delayed gambles (x, p, t) , where $x \in \{5, 10, 15, 20, 30, 50, 80, 95, 96, 97, 98, 99, 100, 120, 150, 245, 246, 247, 248, 249, 250, 300\}$, $p \in \{0.1, 0.3, 0.5, 0.7, 0.9, 1\}$, and $t \in \{0, 3, 7, 14, 30, 60, 90\}$. Clearly, other values can be used depending on the subjects, the domain over which preferences need to be modeled, and other contextual conditions.

Given that the number of candidate pairs is manageable (in our implementation, there are in the order of 500–1,500 candidate pairs for each preference model), we identify the next question by enumeration. In particular, we evaluate each candidate pair of alternatives based on its expected effect on the Hessian of the posterior at its mode, and the candidate pair that maximally increases a norm of this Hessian is chosen as the $(q+1)$ th question.¹⁰

Following the literature on experimental design, we use the absolute value of the determinant as the norm of the Hessian. The Hessian of the posterior likelihood on decision maker i ’s parameters after q questions, computed at \hat{w}_{iq} , is equal to $H_{iq} = \sum_{j=1}^q h(X_{ij}^1, X_{ij}^2, \hat{w}_{iq}, \hat{\delta}) - (\hat{D}^{-1})/2$, where $h(X_{ij}^1, X_{ij}^2, \hat{w}_{iq}, \hat{\delta})$ is the Hessian corresponding to one question. We identify the pair $(X_{i(q+1)}^A, X_{i(q+1)}^B)$ that maximizes the expected value of the norm of the Hessian after question $(q+1)$:

$$p_A |\det(H_{iq} + h(X_{i(q+1)}^A, X_{i(q+1)}^B, \hat{w}_{iq}, \hat{\delta}))| + p_B |\det(H_{iq} + h(X_{i(q+1)}^B, X_{i(q+1)}^A, \hat{w}_{iq}, \hat{\delta}))|, \quad (3)$$

⁹ Some of the amounts are very close (e.g., 95, 96, 97, 98, 99, 100) to estimate the time preferences of very patient subjects more reliably (e.g., someone who would prefer \$98 in three months over \$95 today).

¹⁰ We only consider pairs of gambles in which there is no first-order stochastic dominance, pairs of delayed payments in which there is a trade-off between a sooner-smaller and a larger-later payoff, and pairs of delayed gambles in which none of the gambles in the pair has a payoff that is larger, more likely, and sooner compared to the other.

where

$$p_A = \frac{\exp(\hat{\delta}U(X_{i(q+1)}^A, \hat{w}_{iq}))}{\exp(\hat{\delta}U(X_{i(q+1)}^A, \hat{w}_{iq})) + \exp(\hat{\delta}U(X_{i(q+1)}^B, \hat{w}_{iq}))}$$

is the estimated probability that the decision maker will choose gamble A (computed based on \hat{w}_{iq}), and $H_{iq} + h(X_{i(q+1)}^A, X_{i(q+1)}^B, \hat{w}_{iq}, \hat{\delta})$ is the value of the Hessian after $(q+1)$ questions if gamble A is chosen. This approach is consistent with the Bayesian experimental design literature (e.g., Chaloner and Verdinelli 1995). In a Bayesian decision-theoretic sense, the researcher's utility function is the Hessian of the posterior at its mode. Each question $(q+1)$ is selected sequentially to maximize the expected value of this utility function. As noted by Chaloner and Verdinelli (1995), approximations are typically necessary when dealing with nonlinear designs like ours. Our optimization is approximate to the extent that we approximate $H_{i(q+1)}(\hat{w}_{i(q+1)})$ with $H_{i(q+1)}(\hat{w}_{iq})$, and we compute p_A and p_B based on the point estimate \hat{w}_{iq} instead of the entire posterior after q questions.

In summary, questions are constructed adaptively by performing the following computations between the q th and the $(q+1)$ th question, for each respondent i and for all values of q :

- (i) Update the value of the mode of the posterior distribution, \hat{w}_{iq} .
- (ii) Out of all candidate questions that have not been shown to that respondent, select the one that maximizes the expected value of the determinant of the Hessian of the posterior distribution evaluated at its mode.

To further reduce delays between questions and simplify the use of DEEP, for our online experiment, we computed all possible question paths once and created a large contingency table that indicates which question should be asked following any possible sequence of previous questions and answers. In our implementation with 16 and 20 questions for CPT and QTD, respectively, the number of rows in the table is $2^{16} - 1 = 65,535$ and $2^{20} - 1 = 1,048,575$, respectively, where each row contains a question, the question that precedes it in the path, and the answer to that preceding question that would lead to the question. We provide the first rows of these tables (corresponding to the first five questions) as an illustration in Tables B.1 and B.2 in Appendix B. During the questionnaire, questions are designed by simply looking up the correct values in that table. Our code is available upon request and can be used to generate such tables of questions for any setting.

4. Simulations

Before reporting the results of an online experiment that tested our approach against a benchmark with

real decision makers, we report the results of a set of simulations that investigate how well the proposed approach recovers true parameters. Our initial simulations assess parameter recovery for each of the five preference models considered so far, under different levels of response error and with different numbers of questions. We are particularly interested in the rate of convergence of the estimates as the number of questions increases, and in the accuracy of the estimates. We also report the results of a complementary set of simulations that assess convergence in situations in which the prior is far from the true parameters. We first describe the design of our experiments, followed by the results.

4.1. Simulation Design

Our initial simulations followed a 5×3 full factorial experimental design: we tested the five versions of DEEP corresponding to the five preference models considered above (CPT, sign-dependent CPT, QTD, Ebert–Prelec, and Baucells–Heukamp), under three different levels of response error. In each condition, we tested how our approach is able to recover the true preference parameters of a representative decision maker, w_{true} . For CPT and QTD, we set w_{true} to the mean estimates based on our online experiment (although it is reported later in the paper, the online experiment was conducted before the simulations): $w_{\text{true}} = [0.526; 0.473; 1.682]$ and $[0.925; 0.012]$, respectively for CPT and QTD. Although our online experiment only tested the CPT and QTD versions of our method, we leverage the parallels between Ebert–Prelec and Baucells–Heukamp and CPT and QTD to set w_{true} for all models. For sign-dependent CPT, we introduce arbitrary differences between α^+ and α^- and between σ^+ and σ^- . In particular, we set w_{true} to $[0.626; 0.426; 0.373; 0.573; 1.682]$, $[0.526; 0.012]$, and $[0.526; 0.012; 0.473]$, respectively, for sign-dependent CPT, Ebert–Prelec, and Baucells–Heukamp.

For each model, we simulate the answer to each question according to the logistic probability specified in Equation (1). We parameterize response error as the average proportion of questions that are answered in a way that is inconsistent with the decision maker's true preferences, i.e., the proportion of questions in which the choice probability corresponding to the option chosen by the decision maker is less than 0.5. We vary this proportion, referred to as γ , such that $\gamma = \{0.1, 0.2, 0.3\}$. For each model and each value of γ , we find (through a preliminary set of simulations) the value of the parameter δ that gives rise to the desired average proportion of response errors. For both CPT and QTD, the estimates of δ from our online study are between the values of δ corresponding to $\gamma = 0.2$ and $\gamma = 0.3$ in the simulations. This suggests that the level of response error to be expected in practice is within the range used in our simulations.

For each combination of a model \times level of response error, we simulate $N = 500$ questionnaires each with $Q = 60$ questions (this number of questions is commensurate with those used in previous simulations, e.g., Wang et al. 2010). After each question q , we estimate the parameters based on the answers to the first q questions by finding the mode of the posterior distribution of the parameters, \hat{w}_q . This is done by minimizing the loss function in Equation (2), using $\hat{\delta} = \delta$, and \hat{w}_0 and \hat{D} as described in §2.2.1. This mimics a situation in which the parameters $\hat{\delta}$, \hat{w}_0 , and \hat{D} are set using a pretest or based on the literature (Huber and Zwerina 1996, Sándor and Wedel 2001, Toubia et al. 2007b). We then construct the next question as described in the previous section, by selecting the pair of stimuli that maximizes the expected norm of the Hessian (see Equation (3)).

Although in practice we recommend reestimating all parameters using hierarchical Bayes once the data have been collected (which we do in our online experiment), in our simulations, we focused on the adaptive questionnaire design aspect of our method and did not perform this ex post estimation. Our results are based on the working estimates after each question, \hat{w}_q . We would expect hierarchical Bayes estimates to be more accurate than the ones used here.

4.2. Simulation Results

For each condition (model \times level of response error), we compute the average across replications of the mean absolute percentage error (MAPE) between the estimates and the true parameters after each question. Table 1 reports the average MAPE after 0 (i.e., the MAPE of the mean of the prior distribution, \hat{w}_0), 10, 20, 40, and 60 questions. (To illustrate the impact of the number of parameters on convergence, we order models based on their number of parameters in Table 1 and Figures 1 and 2.)

Our simulations varied three factors: level of response error, number of questions, and preference model. Based on the results, we discuss the impact of each of these three factors in turn.

(i) Comparisons across level of response error (γ). Not surprisingly, we find that within each model, estimates tend to be more accurate when response error is lower. Indeed, the higher the level of response error, the smaller the information provided by each new additional question.

(ii) Comparisons across numbers of questions (q). One should expect estimates to become more accurate as the number of questions becomes larger. This is the case in our simulations, with very few exceptions. In all cases, the estimated parameters converge, on average, toward the true ones.

Table 1 Simulation Results: Mean Absolute Percentage Error Between the Estimates and True Parameters

Model	γ (error)	Number of questions				
		0 (prior)	10	20	40	60
QTD (two parameters)	0.1	0.234	0.027	0.020	0.019	0.019
	0.2	0.234	0.078	0.049	0.041	0.038
	0.3	0.234	0.169	0.131	0.100	0.090
Ebert–Prelec (two parameters)	0.1	0.237	0.062	0.063	0.061	0.061
	0.2	0.237	0.158	0.128	0.104	0.099
	0.3	0.237	0.215	0.215	0.200	0.184
CPT (three parameters)	0.1	0.380	0.175	0.054	0.038	0.036
	0.2	0.380	0.297	0.144	0.089	0.076
	0.3	0.380	0.361	0.282	0.178	0.144
Baucells–Heukamp (three parameters)	0.1	0.388	0.314	0.166	0.065	0.058
	0.2	0.388	0.369	0.211	0.095	0.073
	0.3	0.388	0.489	0.332	0.173	0.131
Sign-dependent CPT (five parameters)	0.1	0.409	0.395	0.365	0.156	0.086
	0.2	0.409	0.405	0.275	0.151	0.104
	0.3	0.409	0.416	0.353	0.243	0.196

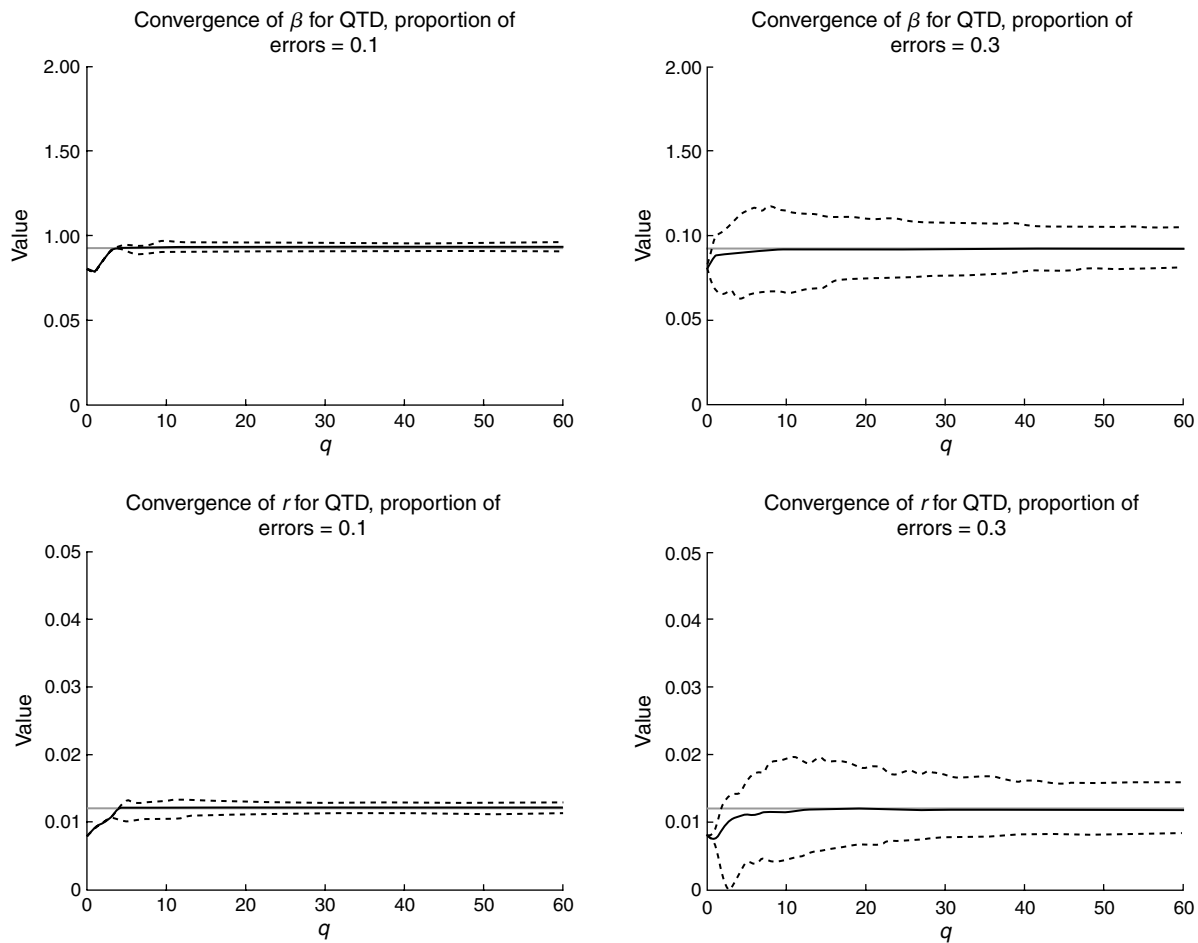
Note. Models are ordered according to their numbers of parameters.

(iii) Comparing across models. Accuracy seems to be driven at least partly by the number of the parameters in the model, as should be expected. Estimates tend to be most accurate and to converge faster for the models that have only two parameters, QTD, and Ebert–Prelec. MAPE tends to be higher and parameter estimates tend to converge slower for the two models that have three parameters, QTD and Baucells–Heukamp, and it is the slowest for the five-parameter model, sign-dependent CPT. This comparison highlights the curse of dimensionality that is typical in preference measurement, and the difficulty of estimating models with many parameters. As noted above, recent adaptive methods such as that of Wang et al. (2010) are limited to two parameters.

For each condition (model \times level of response error), we also plot the evolution of the estimates of each parameter as a function of the number of questions. We plot the average as well as the 95% confidence intervals of the estimates across the 500 replications (the 95% confidence interval is bounded by the 2.5th and 97.5th percentiles of the estimates across replications).¹¹ Because space constraints, we report here only the plots corresponding to CPT and QTD for $\gamma = 0.1$ and 0.3, in Figures 1 and 2, respectively. The plots for the other models give rise to similar conclusions and are available from the authors. The range of the y axis on each plot is the range of possible parameter values, as defined in §2.2.

¹¹ Other researchers have reported the distribution of estimates instead (e.g., Wang et al. 2010). However, this is not practical in our context (three different levels of response error, five different preference models with an average of 3.75 parameters per model).

Figure 1 Simulation Results for QTD: Convergence of the Parameter Estimates



Notes. The solid grey line corresponds to the true value. The solid black line corresponds to the average of the estimate across 500 replications. The dotted lines correspond to the 95% confidence interval across replications.

In conclusion, our simulations suggest that estimates tend to converge to the true values as the number of questions increases. Estimates tend to be more accurate when there is less response error, and when the model being estimated has fewer parameters.

4.3. Robustness to Prior

One feature of the simulations reported above is that the mean of the prior, \hat{w}_0 , is, at least for some parameters, relatively close to the true value used, w_{true} (which, as noted above, was the mean estimate based on our online experiment for CPT and QTD, and based on that for the other models). Although this scenario is realistic and allows studying whether the method is able to improve the accuracy of parameter estimates even when the starting point is already close to the true value, it raises the question of how parameter estimates would converge if the prior and the truth were far from each other. We investigated this issue by rerunning the simulations with different values of \hat{w}_0 . In particular, we ran one set of simulations with $\hat{w}_0 = 2w_{\text{true}}$ and one set with $\hat{w}_0 = 0.5w_{\text{true}}$.

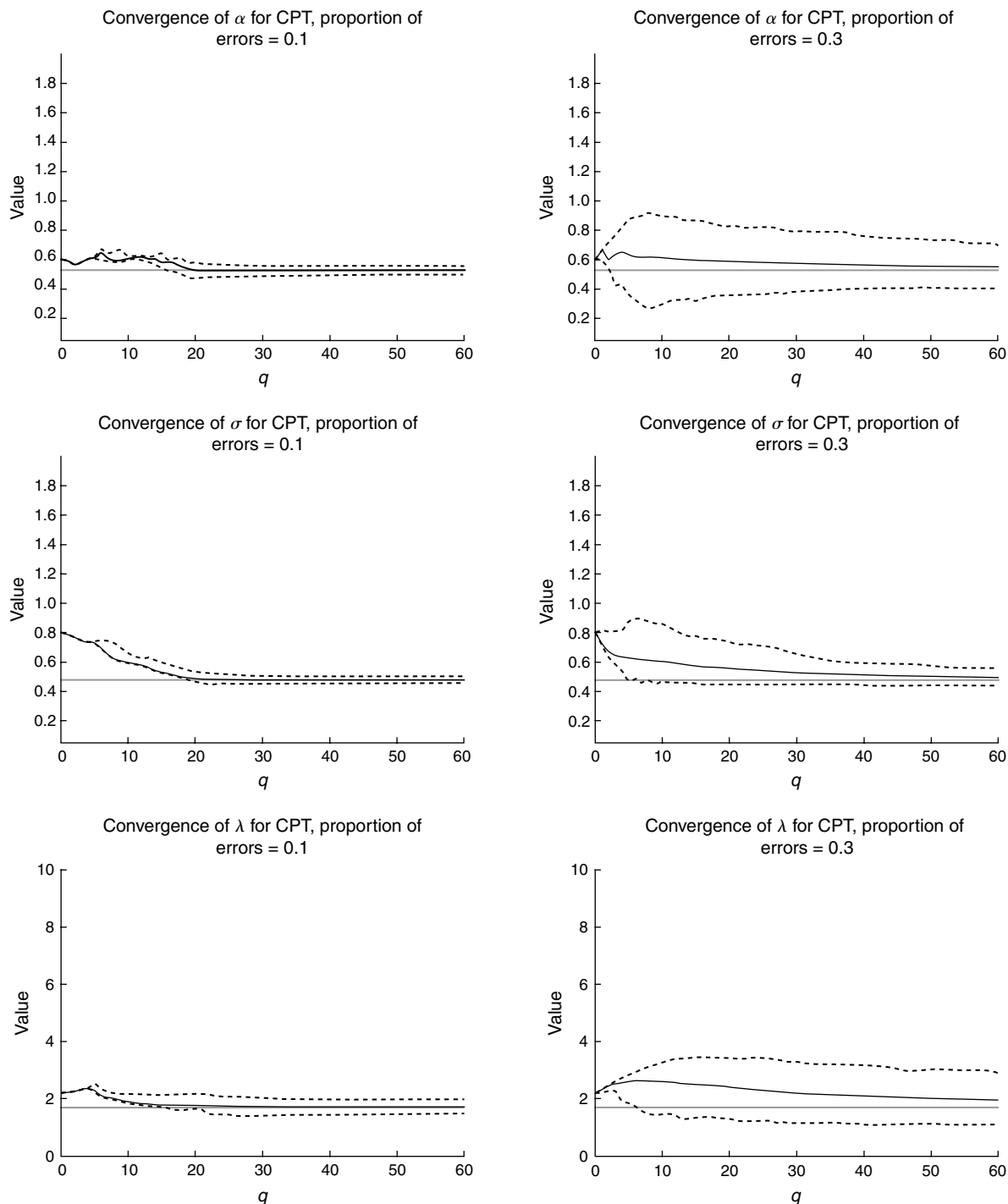
The conclusions are the same as those reached with our initial simulations. We report the convergence plots for CPT and QTD in Figures 3–6. The plots for the other models give rise to similar conclusions and are available from the authors.

5. Online Study

We also tested DEEP using an online study. The purpose of this study was not to conduct a “horse race” with all existing parameter estimation methods in the decision analysis literature, which can be the subject of another extensive empirical project, but rather to examine whether the estimated parameters using DEEP are within the range of values reported in the literature, and to compare its estimation accuracy and time efficiency to those of a common methodology that requires comparable task completion time, which we briefly review next.

5.1. Benchmark

We compared DEEP to a method adapted from Tanaka et al. (2010), because it provides a benchmark

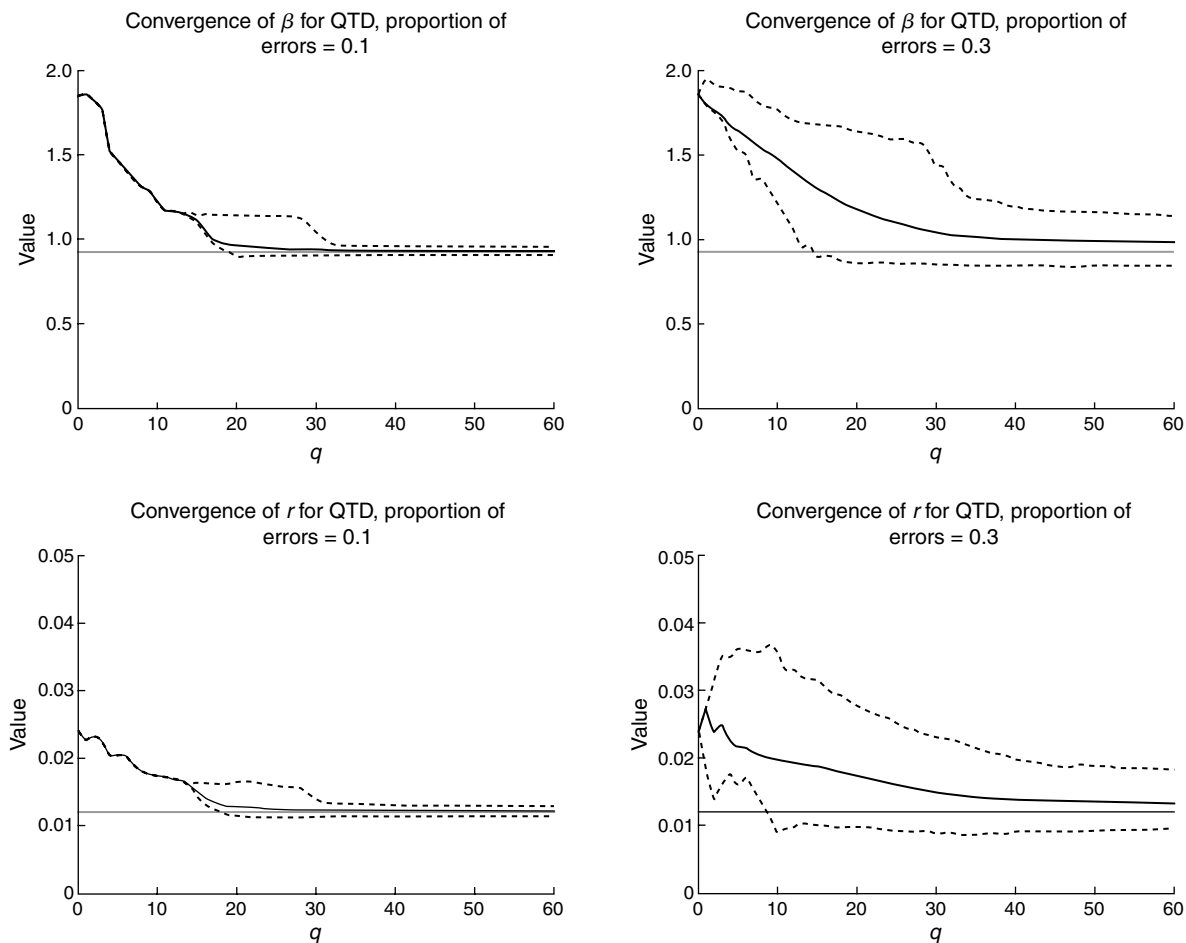
Figure 2 Simulation Results for CPT: Convergence of the Parameter Estimates

Notes. The solid grey line corresponds to the true value. The solid black line corresponds to the average of the estimate across 500 replications. The dotted lines correspond to the 95% confidence interval across replications.

for both risk and time preferences, uses an approach that is common in psychology and economics, is suitable for online administration, and, as also shown below, requires task completion times comparable to DEEP. We focused our empirical tests on the CPT and QTD models, which are the two models on which Tanaka et al. (2010) focus as well.

For the CPT case, subjects in the Benchmark condition were shown three lists of gambles (with 14, 14, and 7 gambles, respectively), identical to Tanaka et al. (2010, Table 2), except that our amounts were in U.S. dollars instead of 1,000 Vietnamese dong. Each list, presented on a single screen, shows one pair of gambles on each row, with the gamble on the right

Figure 3 Simulation Results for QTD When $\hat{w}_0 = 2w_{\text{true}}$



Notes. The solid grey line corresponds to the true value. The solid black line corresponds to the average of the estimate across 500 replications. The dotted lines correspond to the 95% confidence interval across replications.

becoming more and more attractive from row to row (the gambles on the left are constant in the first two lists and less and less attractive from row to row in the third). Subjects were asked to indicate on which row they would start preferring the gamble on the right, if at all. A screenshot of the second series is provided in Figure C.5 in Appendix C. The three series were designed by Tanaka et al. (2010) in such a way that the three CPT parameters can be directly and uniquely calculated from the three switching points. In particular, α and σ are determined jointly from the switching points in the first two lists (see Tanaka et al. 2010, Table A.1). The loss aversion parameter λ is then determined from the switching point in the third list, conditional on the values of the other two parameters elicited from the first two lists.

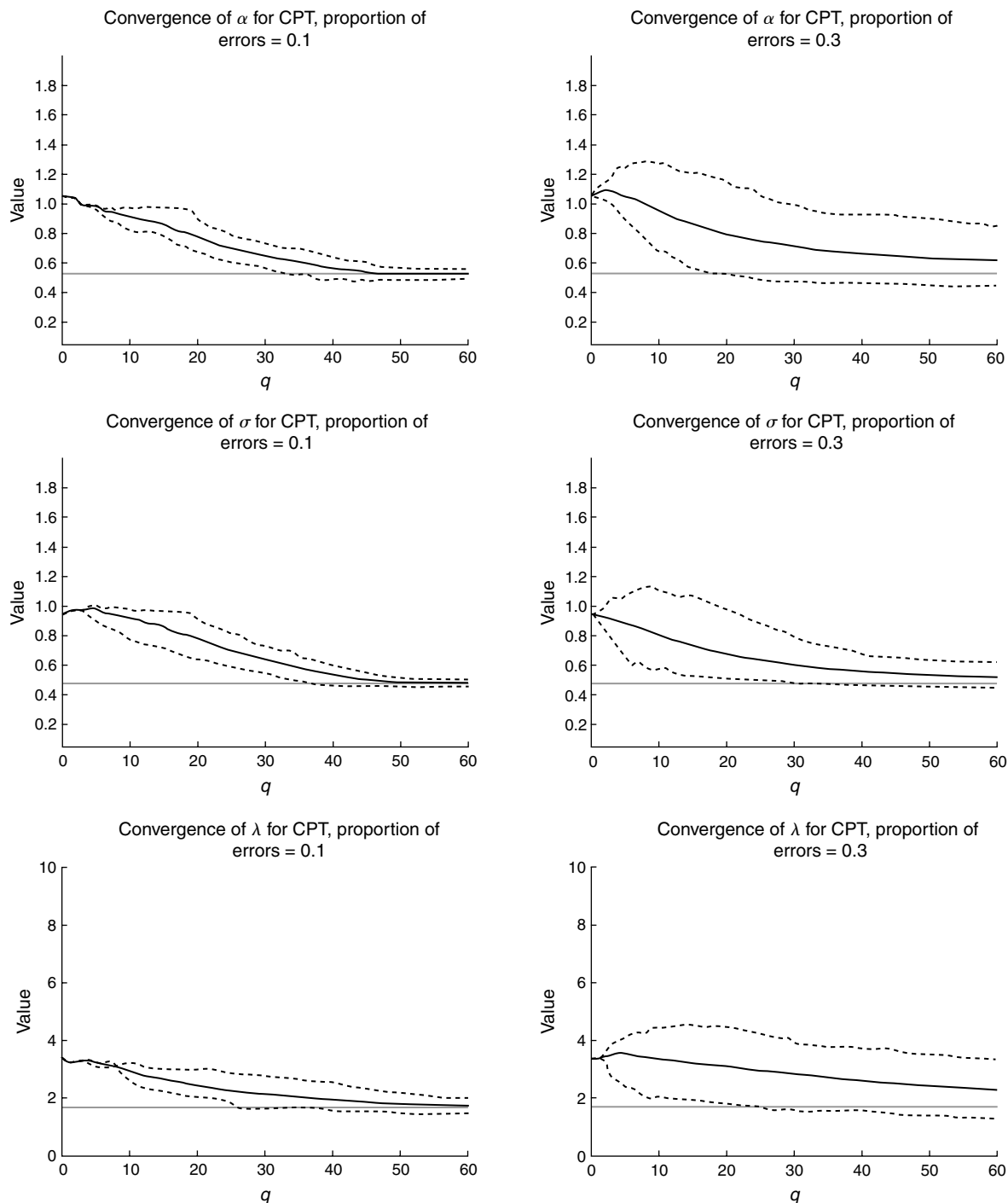
For QTD, our benchmark again followed Tanaka et al. (2010), using U.S. dollars instead of 1,000 dong. In this case, subjects were shown 15 series of five choices between two delayed payment options. In each set, one of the options was fixed while the other

was changing from least to most desirable. An example is shown in Figure C.6 in Appendix C. Effectively, subjects evaluated 75 choices (monotonicity in the subjects' choices within each set was not enforced). Tanaka et al. (2010) do not estimate QTD parameters at the individual level, but instead, use nonlinear least squares to estimate a set of aggregate parameters, allowing for the effect of individual-level covariates on the parameters. To make the comparison with DEEP cleaner, we produced individual-level estimates of the QTD parameters in the Benchmark QTD condition, using the same hierarchical Bayes estimation framework as the one used in the DEEP condition.¹²

5.2. Study Design

We used an online panel of subjects recruited using Amazon's Mechanical Turk who were paid \$2 for their participation, and also could play one of their

¹² Using the estimation method in Tanaka et al. (2010) did not alter our experimental conclusions.

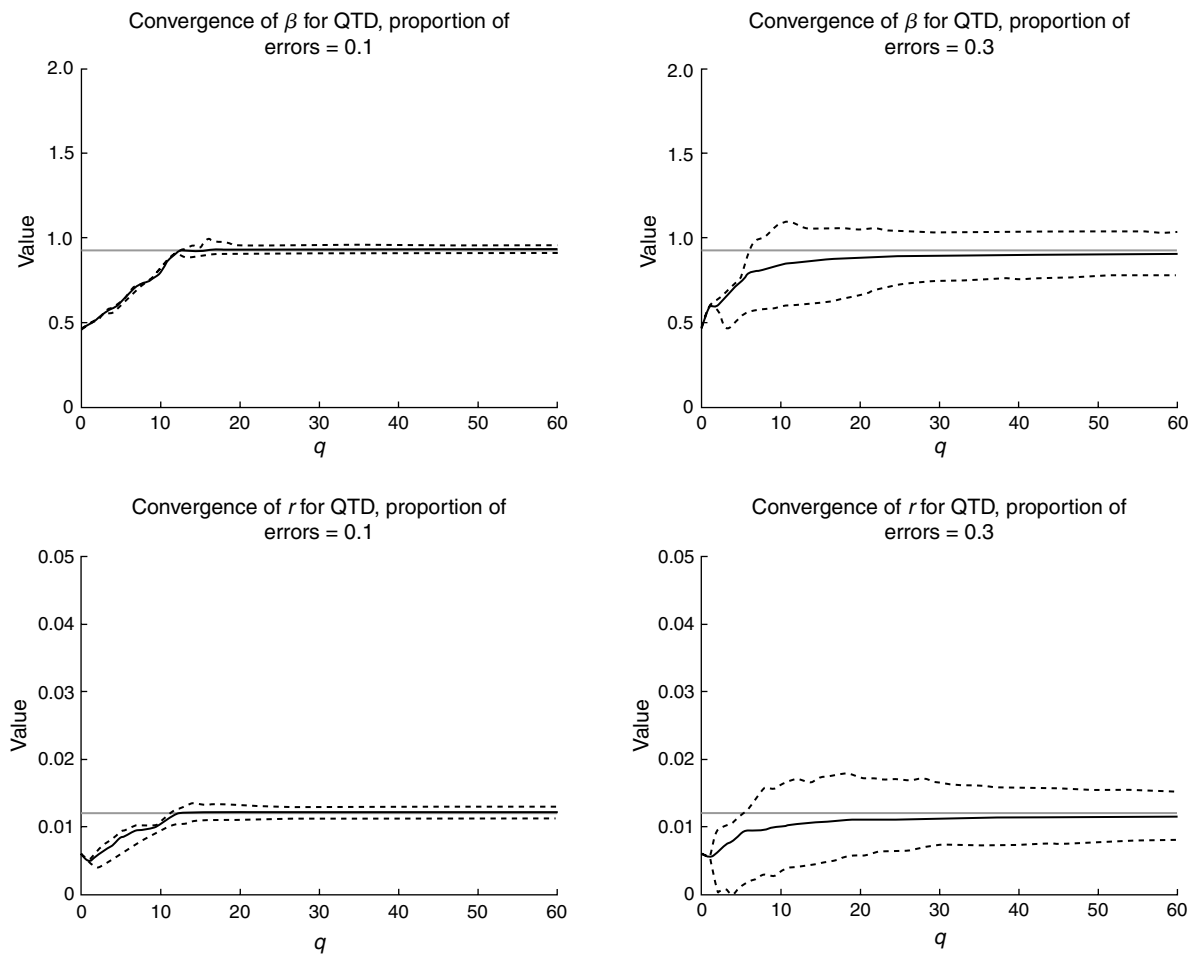
Figure 4 Simulation Results for CPT When $\hat{w}_0 = 2w_{\text{true}}$ 

Notes. The solid grey line corresponds to the true value. The solid black line corresponds to the average of the estimate across 500 replications. The dotted lines correspond to the 95% confidence interval across replications.

(randomly selected) choices for real if they were randomly selected. Data from this pool have been used to replicate many standard decision-making studies, such as the Asian disease problem (Paolacci et al. 2010), and have both interitem and test-retest reliabilities that are equivalent to traditional methods

(Buhrmester et al. 2011). We used a 2×2 between-subject design. Subjects were randomly assigned to either a time preference measurement condition or a risk preference measurement condition, and to DEEP or to the Benchmark method outlined above. For risk preferences (CPT model), we had 137 subjects

Figure 5 Simulation Results for QTD When $\hat{w}_0 = 0.5w_{\text{true}}$



Notes. The solid grey line corresponds to the true value. The solid black line corresponds to the average of the estimate across 500 replications. The dotted lines correspond to the 95% confidence interval across replications.

assigned to DEEP and 133 assigned to the Benchmark. For time preferences (QTD model), we had 150 subjects assigned to DEEP and 146 to the Benchmark.

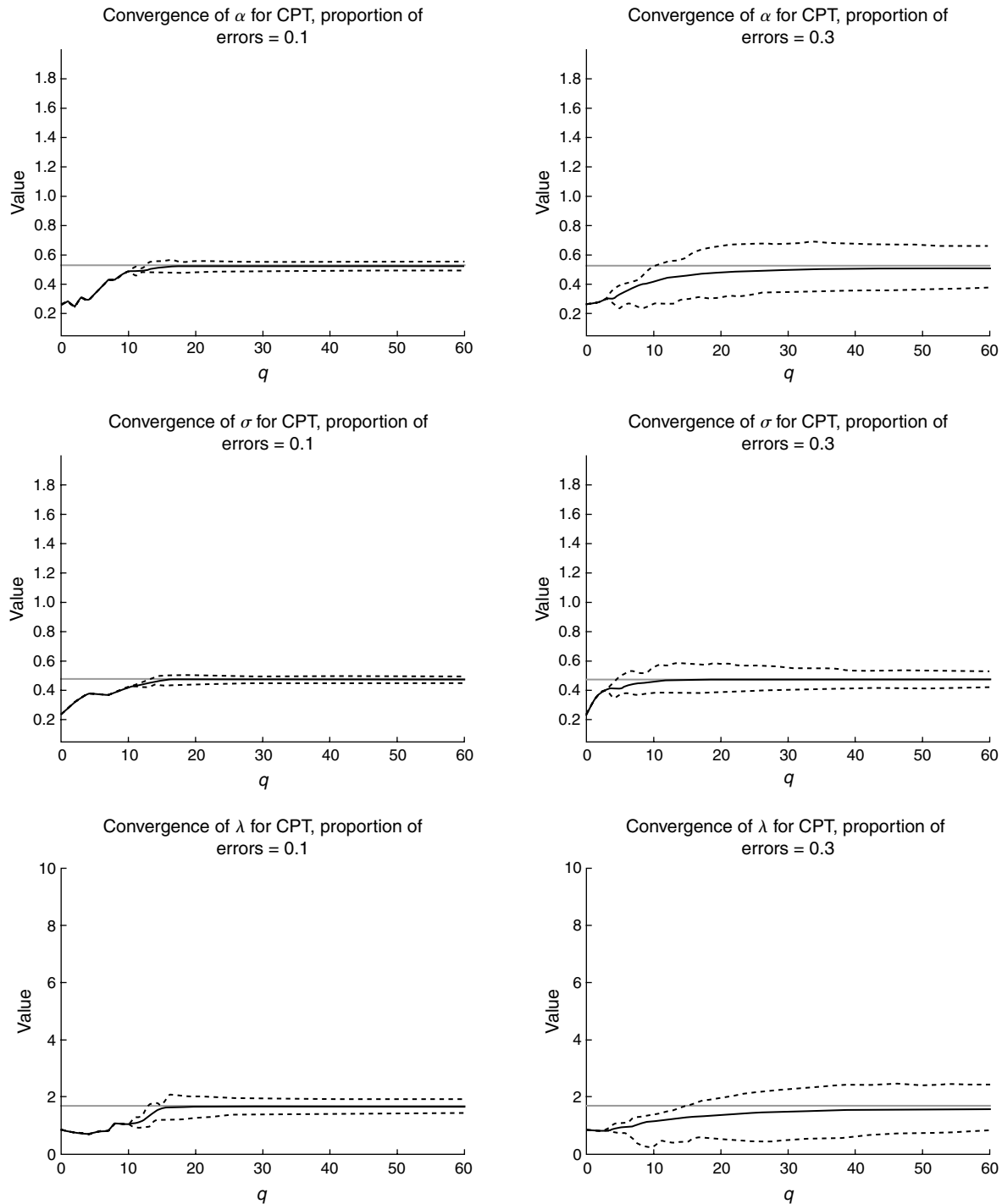
All subjects first saw a welcome page with instructions about the tasks, and they were asked to answer a few simple questions to ensure they understood the subsequent choice questions. For example, for the CPT conditions, subjects were shown a gamble, e.g., $\{\$20, 0.7; \$5\}$, and they were asked questions such as “What is the maximum amount you could win if you played this gamble?” or “Which outcome is most likely?” Subjects who would not answer correctly these comprehension questions returned to the instructions page.

Following these instructions, all subjects were asked to complete an external validity task, which we used to assess the estimation accuracy of the two methods. The external validity task was administered first, so that it would be untainted by, hence comparable between, the two elicitation methods (DEEP and Benchmark).

The external validity task for CPT asked the subjects to indicate their willingness to pay (WTP) for eight gambles.¹³ The gambles for the eight WTP questions for CPT formed a fractional factorial design with $x \in \{\$50, \$100, \$500, \$1,000\}$, $p \in \{0.05, 0.4, 0.6, 0.8\}$, and $y \in \{-\$20, -\$10, \$5, \$20\}$. The set was chosen such that all gambles had an expected value ranging from \$15 to \$80. The list and an example of these WTP questions are shown in Table C.1 and Figure C.1, respectively, in Appendix C.

The external validity task for QTD consisted of eight questions asking the subjects to specify the amount of money that would make them indifferent between a smaller-sooner reward and a larger-later

¹³ To avoid using an external validity task that may be more similar in format to one of the two methods under comparison, we chose a judgment response task, which is different from both (choice-based) elicitation methods. Although there are known inconsistencies between the choice-based and judgment-based elicitation of preferences; namely, preference reversals, these should impact the predictive performance of both DEEP and the Benchmark.

Figure 6 Simulation Results for CPT When $\hat{w}_0 = 0.5w_{\text{true}}$ 

Notes. The solid grey line corresponds to the true value. The solid black line corresponds to the average of the estimate across 500 replications. The dotted lines correspond to the 95% confidence interval across replications.

reward. The format of the task was similar to that in Benhabib et al. (2010). Each subject was asked four acceleration questions (eliciting the amount of money y that would make the subject indifferent between receiving x in t days and y today) followed by four delay questions (eliciting the amount of money y that would make the subject indifferent between receiving

x today and y in t days). For each subject and in each block, the four time-dollar amount combinations (x, t) were drawn randomly without replacement from the nine possible combinations obtained by crossing $(\$10, \$30, \$100)$ with $(3 \text{ days}, 30 \text{ days}, 180 \text{ days})$. An example is shown in Figure C.2 in Appendix C.

After subjects completed the external validity questions, they either completed the DEEP or Benchmark elicitation task for CPT or QTD. For the elicitation of CPT parameters, DEEP asked 16 questions designed according to the proposed adaptive method. A screenshot of one such question is shown in Figure C.3 in Appendix C. For QTD, DEEP asked 20 questions designed according to the proposed adaptive method. A screenshot of one such question is shown in Figure C.4 in Appendix C. In the CPT conditions, we removed all subjects who provided any WTP in the external validity task that was either zero for a gamble that only had positive outcomes or was higher than the maximum amount in the gamble. We were left with 125 subjects for DEEP and 128 for the Benchmark. No subjects were removed in the QTD conditions, to allow for the possibility that some subjects may have a preference for longer delays, a phenomenon known as negative time preference (Loewenstein and Prelec 1991) or future bias (Ashraf et al. 2006, Meier and Sprenger 2010).

In summary, our final data come from the following four conditions:

Condition 1 (DEEP CPT, $N = 125$). Eight WTP questions (external validity task) followed by 16 DEEP CPT questions.

Condition 2 (Benchmark CPT, $N = 128$). Eight WTP questions (external validity task) followed by the three lists of 14, 14, and 7 gambles from the Benchmark CPT method.

Condition 3 (DEEP QTD, $N = 150$). Eight indifference questions (external validity task) followed by 20 DEEP QTD questions;

Condition 4 (Benchmark QTD, $N = 146$). Eight indifference questions (external validity task) followed by 15 sets of five choices from the Benchmark QTD method.

For incentive compatibility, we followed a standard practice in the literature: The subjects were informed at the beginning and reminded throughout the survey that one in every 100 participants would be selected at random and receive an alternative based on the preferences that he or she indicated during the survey. For those selected, one of the questions answered would be selected randomly and the option chosen would be delivered (see Starmer and Sugden 1991 for a treatment of this random lottery procedure).

Adaptive questionnaires raise issues with incentive compatibility, as subjects may potentially misrepresent their true preferences early in the questionnaire to induce more favorable future questions (Harrison 1986). As noted by Wang et al. (2010), this concern may be more theoretical than empirical, especially if no information is provided to subjects about how questions are constructed. Some solutions to this problem have been proposed by these authors

and others. For example, payoffs may be determined based on one question selected ex ante out of all possible questions (unknown to the subjects). If this question was not part of a subject's questionnaire, it would be asked at the end of the questionnaire. Another possible solution is to infer the subjects' responses to such preselected questions based on their responses to the questionnaire. This latter approach was used and validated in the marketing literature by Ding (2007). Designing incentive compatible studies for adaptive questionnaires is a subject of ongoing research, beyond the scope of this paper.

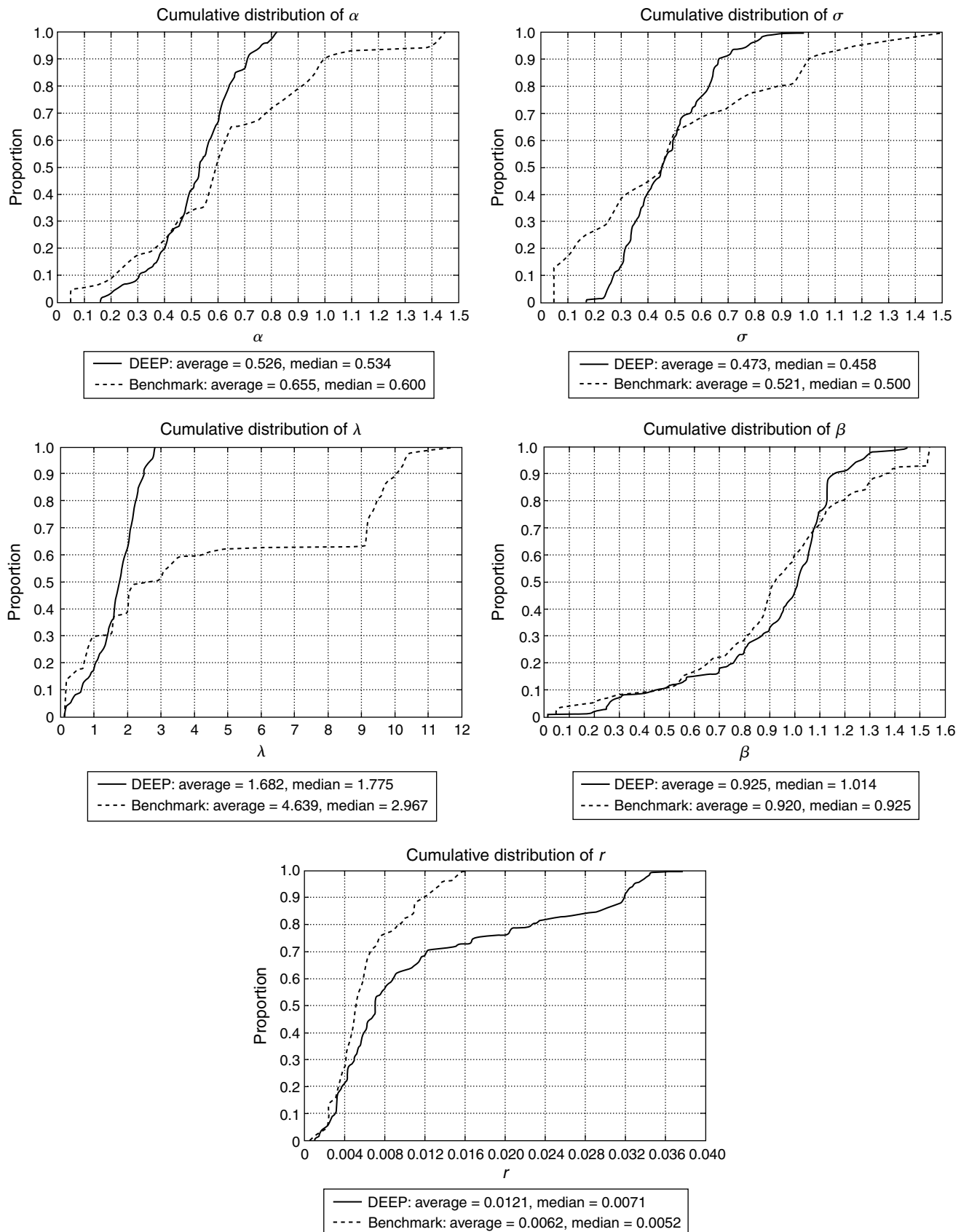
5.3. Experimental Results

5.3.1. Face Validity and Completion Times. Figure 7 reports the cumulative distributions, means, and medians of the estimated parameters for the DEEP and Benchmark conditions. We discuss the CPT parameter estimates first and the QTD ones next.

Although estimates vary across studies with different elicitation methods and subject populations, our estimated CPT parameters are within the range reported in the literature (e.g., Abdellaoui et al. 2007, Gaechter et al. 2011, Liu and Huang 2011, Stott 2006, Tanaka et al. 2010, Tversky and Kahneman 1992, Wu and Gonzalez 1996), providing face validity for the proposed methodology. Comparing DEEP to the Benchmark, the differences in the average probability weighting and loss aversion parameters between the two methods are significant ($p < 0.01$), whereas the difference in the average value function curvature parameter is not. In particular, the Benchmark condition generates higher loss aversion estimates, whereas DEEP estimates are closer to the typical values often observed (around 2). Moreover, the individual parameter estimates obtained by the Benchmark method tend to be more spread out (much more for λ) than those obtained by the DEEP method, as evidenced by the cumulative distributions in Figure 7. Note that differences in the ranges across methods are not inherent to the questionnaire design or estimation methods. The ranges of possible parameter estimates are indeed similar for DEEP versus the Benchmark. Finally, our estimate of δ (the response error parameter in our method) for DEEP is 0.918.

We also report the correlation coefficients between the CPT parameters estimated by the DEEP method.¹⁴ The correlation between σ and λ is negative and statistically significant ($\rho = -0.674$, $p < 0.01$); that between α and σ is also negative and significant, but small ($\rho = -0.215$, $p < 0.02$); that between α and λ is not significant ($\rho = 0.045$). The correlation

¹⁴ None of the correlations between the CPT parameters estimated by the Benchmark method is significant at the $p < 0.05$ level.

Figure 7 Cumulative Distributions, Averages, and Medians of the Parameter Estimates for DEEP and Benchmark

Notes. From top to bottom and left to right: α , σ , λ , β , r . Note that the range of possible parameter estimates are similar for DEEP and the Benchmark.

between σ and λ indicates that individuals with higher σ (value function closer to linear) tend to have lower loss aversion; that is, their value function tends to be uniformly straighter across both the loss and gain domains. This is consistent with the speculation that some people use more rational strategies that might affect both parameters of the value function (Hsee and Rottenstreich 2004). In CPT, an individual's risk attitude is jointly determined by the degree of value function curvature and the degree of probability weighting. The significant, but small negative correlation between σ and α indicates that more curvature in an individual's value function (lower σ) tends to be associated with less probability transformation (higher α). In other words, this suggests that risk attitude tends to be carried more strongly by probability weighting for some individuals, while driven by non-linear valuation of outcomes for others. In a recent study, Qiu and Steiger (2011) found no significant correlation between probability weighting and value function curvature in the gain domain. They inferred from this that it is necessary to have both to adequately capture individuals' risk attitudes. Although our results are different, the small correlation also supports the idea that both elements may be needed to capture the variety of individuals' risk preference types.

Our time preference estimates are in the range obtained from incentive compatible field studies that estimated QTD preferences (Ashraf et al. 2006; Meier and Sprenger 2009, 2010). The average discount rate in the DEEP (respectively, the Benchmark) condition is equivalent to a monthly discount factor of 0.696 (respectively, 0.829). The mean values of the present bias parameter for DEEP and the Benchmark method are very close, but the estimated discount rate from DEEP is higher than that of the Benchmark method (this difference is significant, $p < 0.01$). In addition, we observe a strong negative correlation between β and r in both the DEEP and Benchmark estimates, -0.654 and -0.581 , respectively (both $p < 0.01$); that is, stronger present bias (lower β) is associated with higher discounting of the future. Similar correlations between discount rates and present bias have been reported elsewhere (Meier and Sprenger 2009). Finally, our estimate of δ for DEEP is 0.0525 and our estimate of δ for the Benchmark is 0.111. The fact that the estimate of δ is larger (lower probability of response error) for the Benchmark compared to DEEP is not surprising, because many of the questions asked by the Benchmark are easy to answer (typically the first and/or last questions in each set; see Figure C.6 in Appendix C). The fact that the estimate of δ is much smaller for time preferences compared to risk preferences is not surprising either, because our estimation of time preferences assumed a linear utility

Table 2 Online Study: Completion Time (in Minutes) for Each of the Four Conditions

	DEEP			Benchmark		
	Mean	Median	SD	Mean	Median	SD
Risk preferences (CPT)	2.95	2.78	1.22	3.26	2.83	1.76
Time preferences (QTD)	2.42	2.07	2.21	4.04	3.52	2.06

Note. Bold indicates significantly shorter time at $p < 0.05$ (within a row).

function (no curvature, $\sigma = 1$), while our estimation of risk preferences allowed for concavity in the utility function. As a result, the values of $U(X_{ij}^A, w_i)$ and $U(X_{ij}^B, w_i)$ in Equation (1) are larger for time preferences, and they are multiplied by a smaller value of δ to give rise to choice probabilities that are comparable to the ones obtained in the estimation of risk preferences.

We also report the average time it took subjects to complete the main task, that is, either the DEEP or Benchmark elicitation task for each of the four conditions (excluding the external validity task and the brief introductory task, which were identical for the DEEP and Benchmark conditions). These results are summarized in Table 2. For the elicitation of risk preferences, the DEEP and Benchmark methods take about the same time on average (the difference in means or medians is not significant). For the elicitation of time preferences, DEEP takes significantly less time than the Benchmark method. This difference is significant whether the comparison is done on means or medians ($p < 0.01$).

From the subjects' standpoint, the DEEP method is relatively easy: a few one-shot choice questions (appropriately designed for maximum informativeness). Completion time is often a concern in designing preference assessment experiments, because lengthy questionnaires may cause subjects' responses to become unreliable due to boredom or fatigue and/or may reduce completion rates (Deutschens et al. 2004, Galesic and Bosnjak 2009). The fast completion time offered by DEEP is an attractive feature, particularly for deployment in field studies or online surveys, where subjects are not as "captive" as in a lab.

Finally, we note that, as expected, there were no differences between the durations of the external validity task for the two CPT conditions, nor for the two QTD conditions. All took about two minutes.

5.3.2. External Validity Task: Prediction Accuracy. We now compare DEEP and the Benchmark method on their performance in predicting subjects' responses to the external validity task questions. Based on each subject's estimated preference model parameters, we can calculate predicted responses to the external validity questions for this subject and compare them to the subject's actual responses. For

Table 3 Online Study: Median Absolute Deviation Between the Predicted and Observed Responses Across Eight External Validity Questions

	DEEP	Benchmark
Risk preferences (CPT)	6.56	10.00
Time preferences (QTD)	9.01	9.26

Notes. Bold indicates significantly better at $p < 0.05$ (within a row). We use medians for comparison because the estimated WTP for a few subjects in Condition 2 (CPT, Benchmark) are extremely large—and unrealistic—leading to extremely large mean absolute deviations.

the CPT conditions, we calculate predicted WTP by assuming segregation of the stated payment from the gamble (Thaler 1985). That is, we calculate the WTP such that the disutility of the payment is compensated by the utility of the gamble, according to the CPT model. There is experimental evidence in support of segregation; that is, subjects do not formulate their WTP by mentally subtracting a buying price from the gamble's outcomes and evaluating the gamble net of the price (Casey and Delquie 1995). In the QTD conditions, we calculate the dollar amount that would achieve indifference between the two options, that is, equate their utilities, according to each subject's estimated QTD parameters.

For each subject, we computed the median absolute deviation between the predicted and elicited responses across the eight external validity questions. We then computed the median of these values across the subjects and compared DEEP with the Benchmark both for CPT and QTD.¹⁵ The results are shown in Table 3. DEEP produces significantly more accurate predictions compared to the Benchmark method for CPT ($p < 0.02$). For QTD, there is no significant difference between the prediction accuracy of DEEP and the Benchmark ($p > 0.90$). Note, however, that for QTD, the Benchmark method asked many more questions (presenting 75 choices) and took significantly more time to complete.

Finally, we tested the robustness of these comparisons to the specification of the prior in our hierarchical Bayes estimation. Alternatives to the use of (truncated) normal prior distributions include, for example, uniform distributions, lognormal distributions, mixtures of normal distributions, or highly flexible semiparametric distributions (see, e.g., Ansari and Mela 2003, Kim et al. 2004, Nilsson et al. 2011). We reanalyzed the data from our online experiment using uniform and lognormal priors. The comparisons reported above do not change qualitatively

¹⁵ We use medians for comparison because the estimated WTP for a few subjects in Condition 2 (CPT, Benchmark) are extremely large—and unrealistic—thus grossly inflating the mean. Huge predicted WTP values result from a few subjects in Condition 2 (CPT, Benchmark) having extremely small σ and very low λ .

under these alternative priors (details are available from the authors).

5.3.3. Summary of Results. We conducted an online survey involving more than 500 subjects to compare the proposed DEEP method of preference elicitation with a Benchmark method. DEEP leads to better out-of-sample predictive accuracy for equal time efficiency on the CPT model, and equal predictive accuracy with greater time efficiency on the QTD model. The risk preference (CPT) and time preference (QTD) parameters obtained from DEEP and the Benchmark are different, but all are in the ranges identified in the literature.

6. Discussion

There is a growing interest in relating behavioral decision theories to real-world decisions in areas such as consumer finance (e.g., credit, retirement savings, investment, insurance), health (e.g., nutrition, exercising, substance abuse, medical testing), and others. Success in establishing such links will be highly dependent on our ability to develop time efficient and accurate methods to measure the behavioral characteristics of possibly many economic agents. The proposed methodology has already been used in several such studies, allowing us to further assess its efficiency (time completion as well as face validity of the estimated parameters) in practice and among different subject populations. We briefly review three such studies first and then discuss opportunities for future research.

6.1. Case Studies Using DEEP

We report in Table 4 the mean, median, and standard deviation of the completion times, from three studies that used the proposed methodology. In all studies, completion times were comparable to the ones of our online study above, even though quite different populations were used. In addition to the time necessary to complete the assessment task, both time and risk preference elicitations were preceded by a short introduction to the questions and a test of understanding of

Table 4 Summary of Completion Times (in Minutes) for the Three Case Studies Using DEEP

	Mortgage choice $N = 246$	Social Security $N = 414$	Hormone levels $N = 173$
CPT			
Mean	2.86	4.61	7.42
Median	2.32	3.57	3.54
SD	2.59	5.45	18.12
QTD			
Mean	1.94	2.93	3.27
Median	1.52	2.47	1.86
SD	1.65	1.86	13.33

basic concepts. The median time required to complete the instructions and the elicitation task ranged from 1.5 minutes to 4.1 minutes across studies. Moreover, in all studies, the estimated CPT and QTD parameters were within the range reported in the literature, providing further evidence for the face validity of the proposed methodology.

The first study (Johnson et al. 2011) examined the mortgage decisions of an online panel of 244 homeowners whose average age was 39.6. The purpose of that study was to compare the risk and time preferences (CPT and QTD) of homeowners who owed more on their mortgage than the value of the property to those who did not. A standard titration methodology was also used in that study as a benchmark. An important finding of that study was that, while the standard titration methodology did not reveal any differences between these two kinds of homeowners, DEEP found that those who had negative equity had significantly larger discount rates r as well as present bias (smaller β). This is consistent with the idea that present bias overweighs the immediate value of home possession, and that high discount rates lead to underestimating the longer-run difficulty of meeting payments. These results are consistent with results in other areas of credit (Ashraf et al. 2006; Meier and Sprenger 2009, 2010), providing further face validity to the proposed methodology, and also indicating the potential of DEEP to uncover empirical findings that other methods (e.g., the titration one also used in that study) might not.

A second study (Appelt et al. 2011) examined when potential retirees would start claiming Social Security benefits. The subjects to this online survey were older (mean age = 60.2) and had lower household incomes (median about \$35,000, below the U.S. median), yet completed both the CPT and QTD versions of DEEP within comparable completion times as in the other studies. The decision to retire earlier for a reduced payment or to retire later for a greater payment is a classic intertemporal choice between lower benefits now and larger benefits later. The authors found that the present-bias parameter (β) was related to the decision to claim earlier at a reduced monthly payment, but only for those who currently faced the decision, that is, those for whom retirement was a *now* option.

A third study (Carney et al. 2011) used CPT and QTD estimates provided by the DEEP method to relate measures of current and prenatal hormone levels to risk and time preferences (for related discussions, see Sapienza et al. 2009, Stanton et al. 2011). The subjects were MBA students. Interestingly, although the median completion time was similar to the other studies, there was a group of students who had much longer response times. Self-reports from these MBA students suggested that they were

calculating expected value and discounted utility for the DEEP questions. The authors find, for example, that among male subjects, lower 2D:4D ratio (ratio of the lengths of the second and fourth fingers; a lower ratio is a measure of higher exposure to testosterone in the uterus) is associated with lower values of λ and higher values of σ , that is, a more linear value function.

These three case studies demonstrate that the DEEP methodology can be applied to many populations to produce estimates of time and risk preferences. It seems quite practical for inclusion in online surveys that might involve a large number of subjects. However, in situations where there are few subjects, or where subjects may be available for extended elicitation processes, such as in laboratory settings, other methods might be preferred. As illustrated in these case studies, DEEP does suggest that preference elicitation of relatively complex behavioral models can now be included in unattended online research.

6.2. Discussion and Future Work

We proposed a novel methodology for estimating risk and time preferences. So far, we have applied this method to five preference models (CPT, sign-dependent CPT, QTD, Ebert–Prelec, and Baucells–Heukamp), but the approach could be applied to other models as well. Our approach augments traditional approaches to preference elicitation in decision analysis and bridges the preference measurement literature in marketing with the preference assessment literature in decision theory. The proposed methodology dynamically (i.e., adaptively) optimizes the sequence of questions presented to each subject while modeling response error and leveraging the distribution of the parameters across individuals (heterogeneity). While estimates are available on the fly for each decision maker (based on fixed priors), we recommend reestimating the parameters for all decision makers simultaneously once all the data have been collected, using hierarchical Bayes.

We tested all five versions using simulations and found, as expected, that (i) estimates tend to be more accurate when response error is lower, (ii) estimates tend to be more accurate when the number of questions is larger, and (iii) estimates tend to be more accurate for models with fewer parameters. We also conducted an online study comparing the proposed method to a standard approach used in the literature for the CPT and QTD models. The proposed method either performed significantly better on out-of-sample predictions (for CPT) or took significantly less time to complete (for QTD). This study, as well as three other studies that have used the proposed method, also suggest that the parameters estimated using DEEP have good face validity. The three other studies briefly

reviewed also illustrate how the proposed method may enable researchers to uncover relations between time and risk preferences and other covariates or behaviors. Moreover, the proposed methodology can be deployed by researchers easily with the use of an automatically precomputed table of question paths, as done in our online implementation.

Our simulations, online study, and initial applications indicate the benefits of the DEEP methodology; however, future work is needed to better understand its potential strengths and weaknesses. We close by briefly discussing some issues whose study may both shed light to the limitations of the proposed methodology as well as lead to potential improvements.

One limitation of our approach is that it is myopic, that is, it selects questions one step at a time. Future research may optimize efficiency over the entire set of possible paths (i.e., optimize efficiency over all possible contingency tables such as the ones partially reported in Tables B.1 and B.2 in Appendix B). Although theoretically appealing, such generalization poses great computational challenges. A compromise may be to optimize questions in batches instead of one at a time. However, we note that in a different context (adaptive elicitation of consumer heuristics for product consideration decisions), Dzyabura and Hauser (2011) find that optimizing questions nonmyopically does *not* have a great positive impact on performance.

Another current limitation of our approach is that at each step, the optimal question is identified by enumeration; that is, the expected Hessian is computed for each possible question in a candidate set and the question with the greatest impact on the expected norm of the Hessian is selected. In this paper, choice alternatives were defined by at most three dimensions (e.g., two amounts and one probability), making it possible to construct a set of candidate alternatives of manageable size. Future research may extend our approach to preference models in which the set of possible choice alternatives is too large for enumeration. (Note that the limitation comes from the dimensionality of the alternatives, not the dimensionality of the preference parameters.) One way to address this issue may involve identifying the eigenvector associated with the smallest eigenvalue of the Hessian, and constructing a question based on this eigenvector (as in Abernethy et al. 2008; Toubia et al. 2003, 2004, 2007a, b).

We have focused on the estimation of parameters for several specific models, but the proposed approach can be applied to any preference model, potentially also with modified probability distributions (for the prior and/or the likelihood) and specific question design criteria. The present results suggest that it would be important to further explore the

potential of the proposed methodology in a broader range of behavioral decision-making studies.

More generally, we hope that this work not only enables more work using models of preferences in many more settings, but also that it encourages a dialog between those interested in the psychology of preference and those interested in preference measurement and its application. Traditionally, the appeal of adaptive methods has been statistical efficiency and minimization of potentially expensive subjects' time, making studies of large or online populations practically more feasible. However, there are reasons to suspect that adaptive methods are attractive because of other specific characteristics worth studying in the future. First, by limiting the number of questions that are posed to the subject, adaptive methods may minimize the cognitive resources required to assess preferences. The possibility that resource depletion occurs with an increased number of questions seems likely (Vohs et al. 2008), and recent evidence suggests that some context effects, such as selection of a default option, increase with depletion (Levav et al. 2010). Thus, by focusing attention on questions that are most (statistically) informative, adaptive methods might produce estimates that are less contaminated by context effects in addition to being less influenced by the random error produced by fatigue. A second possible advantage of adaptive methods is that they focus more quickly on the set of questions that are most relevant to portraying the decision maker's preferences. Although the algorithm is designed to decrease our uncertainty on the decision maker's preference parameters, from the decision maker's perspective, it may be seen as eliminating less relevant questions. This may also limit the possibility of range effects due to irrelevant extreme values that might be presented to the decision maker. Future work may attempt to better understand the characteristics of the elicitation methods that simultaneously minimize effort, minimize error due to stochastic sources such as inattention, and at the same time, maximize the predictive validity of the measures. Such work would inform three questions that lie at the basis of our endeavor: What are the right adaptive questions? What is the right model of noise? And, finally, what is the right model of preferences?

Acknowledgments

The authors thank Jacob Abernethy and Alain Lemaire for their support with the development of the online platform used in this paper. They also thank the staff at the Center for Decision Sciences at Columbia University, in particular, Cindy Kim, Ye Li, and Jon Westfall, for their support in validating, testing, and modifying this platform. Support is acknowledged from the National Institute on Aging [Grant 3R01AG027934-04].

Appendix A. MCMC Algorithm for Estimating the CPT and QTD Parameters

We consider here the more general setup that allows capturing the effect of covariates on the parameters through the prior distribution on w_i . In particular, we consider the following prior distribution on w_i (see, e.g., Allenby and Ginter 1995, Lenk et al. 1996): $w_i \sim \text{TN}(\Theta z_i, D)$, where z_i is a set of covariates for subject i (including an intercept) and Θ is a matrix capturing the relationship between these covariates and the mean of the first-stage prior (this matrix is estimated). For example, if the covariates are age (in years) and gender (binary equal to 1 for male and 0 for female), then for CPT, the 3×3 dimensional matrix Θ is

$$\Theta = \begin{bmatrix} \theta_{1,1} & \theta_{1,2} & \theta_{1,3} \\ \theta_{2,1} & \theta_{2,2} & \theta_{2,3} \\ \theta_{3,1} & \theta_{3,2} & \theta_{3,3} \end{bmatrix},$$

where $\theta_{1,1}$, $\theta_{2,1}$, and $\theta_{3,1}$ are the intercept parameters for α , σ , and λ , respectively; $\theta_{1,2}$, $\theta_{2,2}$, and $\theta_{3,2}$ capture the effect of age on the parameters α , σ , and λ , respectively; and $\theta_{1,3}$, $\theta_{2,3}$, and $\theta_{3,3}$ capture the effect of gender on the parameters α , σ , and λ , respectively. In this case, the second-stage prior is defined on Θ and D instead of w_0 and D . This specification allows us to capture the effects of covariates on the parameters directly, in an integrated model estimated in one step, instead of using a two-step approach of estimating the parameters first and then regressing them on covariates. Note that the case without covariates is a special case of this formulation, in which the covariates are limited to an intercept (the vector w_0 corresponds to the first column of Θ). Therefore, we describe below the estimation procedure with covariates, as it nests the formulation without covariates. In our experiment, we did not use any covariate to make the comparison with the benchmark cleaner.

MCMC draws successively each parameter from its posterior distribution conditioning on the data and other parameters. Each parameter is drawn once in each iteration. The resulting Markov chain has the posterior distribution as its equilibrium distribution. In our online study, for QTD, we use 10,000 iterations as “burn-in” (i.e., these iterations allow the Markov chain to converge to its equilibrium distribution and are not saved), followed by 40,000 iterations in

which 1 iteration is saved every 10 iterations; for CPT, we use 50,000 iterations as burn-in, followed by 50,000 more iterations (convergence tends to be slower). The outcome is a set of draws from the posterior distribution. We now describe how each parameter is updated at each iteration.

(i) Update of D : This matrix is drawn directly from its conditional posterior distribution

$$P(D | \text{rest}, \text{data}) \sim \text{IW}(\eta_0 + I, \eta_0 \Delta_0 + \sum_{i=1}^I (w_i - \Theta z_i)(w_i - \Theta z_i)^T),$$

We use typical parameter values for the inverse Wishart prior on D : $\eta_0 = n_{\text{par}} + 3$ and $\Delta_0 = 0.1I$ for CPT and

$$\Delta_0 = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.0001 \end{bmatrix}$$

for QTD (to account for the fact that the second parameter is on a smaller scale compared to the others), where n_{par} is the number of parameters in the model.

(ii) Update of $\{w_i\}$: We use a Metropolis–Hastings algorithm, with a normal random walk proposal density (with a jump size adapted to keep the acceptance ratio around 30%). Constraints on the parameters are enforced with rejection sampling (Allenby et al. 1995). For each w_i , the acceptance ratio is obtained from

$$\prod_j \frac{\exp(\delta U(X_{ij}^1, w_i))}{\exp(\delta U(X_{ij}^1, w_i)) + \exp(\delta U(X_{ij}^2, w_i))} \cdot \exp\left(-\frac{1}{2}(w_i - \Theta z_i)^T D^{-1}(w_i - \Theta z_i)\right).$$

(iii) Update of Θ : This matrix is drawn directly from its conditional posterior distribution: $P(\text{Vec}(\Theta) | \text{rest}, \text{data}) \sim N(V(Z^T \otimes D^{-1})\text{Vec}(W), V)$, where $V = ((Z^T Z) \otimes D^{-1})^{-1}$, Z is the matrix of covariates z_i 's (one row per decision maker), W is the matrix of w_i 's (one row per decision maker), $\text{Vec}(X)$ is the column vector obtained by stacking the columns of a matrix X , and \otimes is the Kronecker product.

(iv) Update of δ : We use a Metropolis–Hastings algorithm with a normal random walk proposal density (with variance 10^{-5}). The acceptance ratio is obtained from

$$\prod_i \prod_j \frac{\exp(\delta U(X_{ij}^1, w_i))}{\exp(\delta U(X_{ij}^1, w_i)) + \exp(\delta U(X_{ij}^2, w_i))}.$$

Appendix B

Table B.1 Subset of the Contingency Table for CPT (First Five Questions)

Question index	Previous question	Previous answer	x^A	p^A	y^A	x^B	p^B	y^B
0	N/A	N/A	1	0.9	5	100	0.3	−20
1	0	1	1	0.7	−15	100	0.3	−20
65,536	0	2	1	0.5	−5	40	0.3	−15
2	1	1	1	0.5	−10	100	0.3	−20
32,769	1	2	1	0.5	−5	100	0.3	−20
65,537	65,536	1	40	0.3	−5	100	0.3	−20
98,304	65,536	2	40	0.9	5	100	0.3	−20
3	2	1	1	0.7	−15	40	0.5	−20
16,386	2	2	30	0.1	−10	40	0.3	−15
32,770	32,769	1	1	0.9	5	30	0.3	−5
49,153	32,769	2	1	0.3	10	40	0.3	−5

Table B.1 (Continued)

Question index	Previous question	Previous answer	x^A	p^A	y^A	x^B	p^B	y^B
65,538	65,537	1	1	0.1	5	100	0.3	−20
81,921	65,537	2	1	0.7	−15	40	0.1	−10
98,305	98,304	1	1	0.9	5	40	0.3	−15
114,690	98,304	2	1	0.7	−20	30	0.1	−20
4	3	1	30	0.1	−10	100	0.3	−20
8,195	3	2	40	0.5	−20	1,000	0.1	−5
16,387	16,386	1	1	0.9	5	30	0.3	−5
24,578	16,386	2	30	0.3	−10	100	0.3	−20
32,771	32,770	1	30	0.3	−10	100	0.3	−20
40,962	32,770	2	1	0.3	10	40	0.3	−5
49,154	49,153	1	1	0.9	30	100	0.3	−20
57,345	49,153	2	30	0.1	−10	40	0.3	−15
65,539	65,538	1	1	0.1	5	30	0.3	−5
73,730	65,538	2	1	0.3	10	100	0.3	−20
81,922	81,921	1	30	0.5	−5	100	0.3	−20
90,113	81,921	2	1	0.7	−15	30	0.1	−10
98,306	98,305	1	1	0.3	10	40	0.5	−20
10,6500	98,305	2	30	0.3	−10	40	0.3	−15
114,690	114,690	1	1	0.1	5	40	0.1	−10
122,880	114,690	2	1	0.5	−10	30	0.1	−20

Note. Each row contains the question index, the index of the question that precedes it in the path, the answer to that preceding question that leads to the question, and the two alternatives in the question.

Table B.2 Subset of the Contingency Table for QTD (First Five Questions)

Question index	Previous question	Previous answer	x^A	t^A	x^B	t^B
0	N/A	N/A	95	0	250	90
1	0	1	96	0	300	14
524,288	0	2	248	0	250	90
2	1	1	10	3	250	90
262,145	1	2	95	0	250	30
524,289	524,288	1	250	0	300	7
786,432	524,288	2	249	0	250	30
3	2	1	5	0	300	7
131,074	2	2	20	0	250	90
262,146	262,145	1	150	0	300	14
393,217	262,145	2	250	0	300	7
524,290	524,289	1	150	0	250	90
655,361	524,289	2	150	3	250	90
786,433	786,432	1	249	0	250	90
917,504	786,432	2	249	0	250	90
4	3	1	10	0	300	7
65,539	3	2	50	0	300	7
131,075	131,074	1	10	0	250	90
196,610	131,074	2	50	0	300	7
262,147	262,146	1	150	0	300	7
327,682	262,146	2	245	0	300	7
393,218	393,217	1	80	7	250	90
458,753	393,217	2	96	0	250	60
524,291	524,290	1	150	0	300	7
589,826	524,290	2	245	0	300	7
655,362	655,361	1	250	0	300	14
720,897	655,361	2	245	0	250	30
786,434	786,433	1	248	0	250	30
851,969	786,433	2	248	0	250	30
917,505	917,504	1	248	0	250	30
983,040	917,504	2	247	0	250	90

Note. Each row contains the question index, the index of the question that precedes it in the path, the answer to that preceding question that leads to the question, and the two alternatives in the question.

Appendix C

Table C.1 The Eight Gambles Used in the External Validity Task for CPT

Gamble	Outcome 1(\$)	Probability 1	Outcome 2 (\$)	Probability 2
1	500	0.05	−10	0.95
2	50	0.80	20	0.20
3	100	0.60	5	0.40
4	50	0.40	20	0.60
5	1,000	0.05	5	0.95
6	100	0.40	−10	0.60
7	100	0.80	−20	0.20
8	50	0.60	−20	0.40

Figure C.1 Example of External Validity Task Question for CPT

Survey

Please consider the gamble below. State the maximum that you would be willing to pay to play this gamble. This is the maximum amount that you would pay to be able to play this gamble. If it was offered to you at any higher price, you would not play the gamble. If you really would rather not play the gamble under any circumstances, please put in 0.

Gamble

5% Chance to Win \$500
95% Chance to Lose \$10

The most I would be willing to pay to play this gamble is (please enter a dollar amount):

Submit

SECTION #3 OUT OF 5

Questions Completed
1/8

Figure C.2 Example of External Validity Task Question for QTD

Survey

Please consider the two options below. One option involves receiving some amount of money in the future, and the other involves receiving some amount of money today. While the amount of money received in the future is specified, the other amount is left unspecified. Please enter the amount that would make you indifferent between the two options.

What amount of money would make you indifferent between these two options (please enter a dollar amount in the blank space)

Option A

Option B

Receive \$ today

Receive \$30 in 6 months

Submit

More Information on Prices

SECTION #2 OUT OF 4

Questions Completed
1/4

Figure C.3 Example of DEEP Question for CPT

RiskSense Survey

Please consider the two gambles below. Which of these two gambles would you rather play?

Option A

Option B

50% Chance to Win \$100
50% Chance to Lose \$15

90% Chance to Win \$40
10% Chance to Win \$10

Choose A

Choose B

More Information on Prices

Questions Completed
3/16

Downloaded from informs.org by [155.246.103.35] on 06 April 2017, at 12:21 . For personal use only, all rights reserved.

RIGHTSLINK
Copyright Clearance Center

Figure C.4 Example of DEEP Question for QTD

RiskSense Survey

Please consider the two options below. Which of these two options do you find more attractive?

Option A	Option B
Receive \$250 today	Receive \$300 in 1 week
<input type="button" value="Choose A"/>	<input type="button" value="Choose B"/>

More Information on Prizes

Questions Completed

5/20

Figure C.5 Example of Benchmark Question for CPT

Survey

Please take a look at the table below. You will notice that Option B gets more and more attractive as you go down the table. At what point are you indifferent between options A and B? Please select the first row at which you would start preferring Option B (so for that row as well as the rows below it you prefer B, and for the rows above you prefer A).

Option A	Option B
Always Prefer Option A	<input type="radio"/>
90% Chance to Win \$40 10% Chance to Win \$30	70% Chance to Win \$54 30% Chance to Win \$5
90% Chance to Win \$40 10% Chance to Win \$30	70% Chance to Win \$56 30% Chance to Win \$5
90% Chance to Win \$40 10% Chance to Win \$30	70% Chance to Win \$58 30% Chance to Win \$5
90% Chance to Win \$40 10% Chance to Win \$30	70% Chance to Win \$60 30% Chance to Win \$5
90% Chance to Win \$40 10% Chance to Win \$30	70% Chance to Win \$62 30% Chance to Win \$5
90% Chance to Win \$40 10% Chance to Win \$30	70% Chance to Win \$65 30% Chance to Win \$5
90% Chance to Win \$40 10% Chance to Win \$30	70% Chance to Win \$68 30% Chance to Win \$5
90% Chance to Win \$40 10% Chance to Win \$30	70% Chance to Win \$72 30% Chance to Win \$5
90% Chance to Win \$40 10% Chance to Win \$30	70% Chance to Win \$77 30% Chance to Win \$5
90% Chance to Win \$40 10% Chance to Win \$30	70% Chance to Win \$83 30% Chance to Win \$5
90% Chance to Win \$40 10% Chance to Win \$30	70% Chance to Win \$90 30% Chance to Win \$5
90% Chance to Win \$40 10% Chance to Win \$30	70% Chance to Win \$100 30% Chance to Win \$5
90% Chance to Win \$40 10% Chance to Win \$30	70% Chance to Win \$110 30% Chance to Win \$5
90% Chance to Win \$40 10% Chance to Win \$30	70% Chance to Win \$130 30% Chance to Win \$5
Always Prefer Option B	<input type="radio"/>

SECTION #2 OUT OF 2

Questions Completed

2/3

Figure C.6 Example of Benchmark Question for QTD

Survey

Please take a look at the table below. Notice that as you go down the table one option is fixed while the other changes. In each row of the table, please indicate your preference between option A and option B.

Option A	Option B
Receive \$120 in 1 week	Receive \$20 today
Receive \$120 in 1 week	Receive \$40 today
Receive \$120 in 1 week	Receive \$60 today
Receive \$120 in 1 week	Receive \$80 today
Receive \$120 in 1 week	Receive \$100 today

SECTION #2 OUT OF 2

Questions Completed

1/15

References

Abdellaoui M, Bleichrodt H, L'Haridon O (2008) A tractable method to measure utility and loss aversion under prospect theory. *J. Risk Uncertainty* 36(3):245–266.

Abdellaoui M, Bleichrodt H, Paraschiv C (2007) Loss aversion under prospect theory: A parameter-free measurement. *Management Sci.* 53(10):1659–1674.

Abernethy J, Evgeniou T, Toubia O, Vert J-P (2008) Eliciting consumer preferences using robust adaptive choice questionnaires. *IEEE Trans. Knowledge Data Engrg.* 20(2):145–155.

Allenby GM, Ginter JL (1995) Using extremes to design products and segment markets. *J. Marketing Res.* 32(4):392–403.

Allenby GM, Rossi PE (1999) Marketing models of consumer heterogeneity. *J. Econometrics* 89(1–2):57–78.

Allenby GM, Arora N, Ginter JL (1995) Incorporating prior knowledge into the analysis of conjoint studies. *J. Marketing Res.* 32(2):152–162.

Andrews RL, Ainslie A, Currim IS (2002a) An empirical comparison of logit choice models with discrete versus continuous representations of heterogeneity. *J. Marketing Res.* 39(4):479–487.

Andrews RL, Ansari A, Currim IS (2002b) Hierarchical Bayes versus finite mixture conjoint analysis models: A comparison of fit, prediction, and partworth recovery. *J. Marketing Res.* 39(1):87–98.

- Angeletos G-M, Laibson D, Repetto A, Tobacman J, Weinberg S (2001) The hyperbolic consumption model: Calibration, simulation, and empirical evaluation. *J. Econom. Perspect.* 15(3):47–68.
- Ansari A, Mela CF (2003) E-customization. *J. Marketing Res.* 40(2):131–145.
- Appelt KC, Johnson EJ, Knoll MAZ, Westfall JE (2011) Time to retire: Why Americans claim benefits early and how to encourage them to delay. Working paper, Graduate School of Business, Columbia University, New York.
- Ashraf N, Karlan D, Yin W (2006) Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines. *Quart. J. Econom.* 40(1):635–72.
- Barberis N, Huang M, Santos T (2001) Prospect theory and asset prices. *Quart. J. Econom.* 116(1):1–53.
- Baucells M, Heukamp FH (2012) Probability and time trade-off. *Management Sci.* 58(4):831–842.
- Benhabib J, Bisin A, Schotter A (2010) Present-bias, quasi-hyperbolic discounting, and fixed costs. *Games Econom. Behav.* 69(2):205–223.
- Bruhin A, Fehr-Duda H, Epper T (2010) Risk and rationality: Uncovering heterogeneity in probability distortion. *Econometrica* 78(4):1375–1412.
- Buhrmester M, Kwang T, Gosling SD (2011) Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspect. Psych. Sci.* 6(1):3–5.
- Camerer CF, Loewenstein G, Rabin M (2003) *Advances in Behavioral Economics* (Princeton University Press, Princeton, NJ).
- Carney D, Toubia O, Johnson EJ (2011) Stable and transient influences of hormone levels on time and risk preferences. Working paper, Columbia Business School, Columbia University, New York.
- Casey JT, Delqu   P (1995) Stated vs. implicit willingness to pay under risk. *Organ. Behav. Human Decision Processes* 61(2):123–137.
- Cavagnaro DR, Gonzalez R, Myung JI, Pitt MA (2012) Optimal decision stimuli for risky choice experiments: An adaptive approach. Working paper, Ohio State University, Columbus.
- Chaloner K, Verdinelli I (1995) Bayesian experimental design: A review. *Statist. Sci.* 10(3):273–304.
- De Groot M (1970) *Optimal Statistical Decisions* (McGraw-Hill, New York).
- Deutskens E, De Ruyter K, Wetzels M, Paul O (2004) Response rate and response quality of Internet-based surveys: An experimental study. *Marketing Lett.* 15(1):21–36.
- Ding M (2007) An incentive-aligned mechanism for conjoint analysis. *J. Marketing Res.* 44(2):214–223.
- Dzyabura D, Hauser JR (2011) Active machine learning for consideration heuristics. *Marketing Sci.* 30(5):801–819.
- Ebert JE, Prelec D (2007) The fragility of time: Time-insensitivity and valuation of the near and far future. *Management Sci.* 53(9):1423–1438.
- Evgeniou T, Pontil M, Toubia O (2007) A convex optimization approach to modeling consumer heterogeneity in conjoint estimation. *Marketing Sci.* 26(6).
- Ford I, Kitsos CP, Titterton DM (1989) Recent advances in nonlinear experimental designs. *Technometrics* 31(1):49–60.
- Frederick S, Loewenstein G, O'Donoghue T (2002) Time discounting and time preference: A critical review. *J. Econom. Literature* 40(2):351–401.
- Gaechter S, Johnson EJ, Herrmann A (2011) Measuring individual-level loss aversion using simple experiments. Working paper, Graduate School of Business, Columbia University, New York.
- Galesic M, Bosnjak M (2009) Effects of questionnaire length on participation and indicators of response quality in a Web survey. *Public Opinion Quart.* 73(2):349–360.
- Green PE, Rao VR (1972) Configuration synthesis in multidimensional scaling. *J. Marketing Res.* 9(1):65–68.
- Greene W (2000) *Econometric Analysis* (Prentice-Hall, Upper Saddle River, NJ).
- Harrison GW (1986) An experimental test for risk aversion. *Econom. Lett.* 21(1):7–11.
- Harrison GW, Lau MI, Williams MB (2002) Estimating individual discount rates in Denmark: A field experiment. *Amer. Econom. Rev.* 92(5):1606–1617.
- Hey J, Orme C (1994) Investigating generalizations of expected utility theory using experimental data. *Econometrica* 62(6):1291–1326.
- Hsee CK, Rottenstreich Y (2004) Music, pandas, and muggers: On the affective psychology of value. *J. Experiment. Psych.* 133(1):23–30.
- Huber J, Zwerina K (1996) The importance of utility balance in efficient choice designs. *J. Marketing Res.* 33(3):307–317.
- Jarnebrant P, Toubia O, Johnson E (2009) The silver lining effect: Formal analysis and experiments. *Management Sci.* 55(11):1832–1841.
- Johnson EJ, Atlas SA, Payne JW (2011) Time preferences, mortgage choice, and strategic default. Working paper, Columbia Business School, Columbia University, New York.
- Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–292.
- Kamakura WA, Russell GJ (1989) A probabilistic choice model for market segmentation and elasticity structure. *J. Marketing Res.* 26(4):379–390.
- Kim JG, Menzefricke U, Feinberg FM (2004) Assessing heterogeneity in discrete choice models using a Dirichlet process prior. *Rev. Marketing Sci.* 2(1):1–39.
- Kuhfeld WF, Tobias RD, Garratt M (1994) Efficient experimental design with marketing applications. *J. Marketing Res.* 31(4):545–557.
- Laibson D (1997) Golden eggs and hyperbolic discounting. *Quart. J. Econom.* 112(2):443–477.
- Laskey KB, Fischer GW (1987) Estimating utility functions in the presence of response error. *Management Sci.* 33(8):965–980.
- Lenk PJ, DeSarbo WS, Green PE, Young MR (1996) Hierarchical Bayes conjoint analysis: Recovery of partworth heterogeneity from reduced experimental designs. *Marketing Sci.* 15(2):173–191.
- Levav J, Heitmann M, Herrmann A, Iyengar SS (2010) Order in product customization decisions: Evidence from field experiments. *J. Political Econom.* 118(2):274–299.
- Liu EM, Huang J (2011) Risk preferences and pesticide use by cotton farmers in China. Working paper, University of Houston, Houston.
- Loewenstein G, Prelec D (1991) Negative time preference. *Amer. Econom. Rev.* 81(2):347–352.
- Luce RD (1958) A probabilistic theory of utility. *Econometrica* 26(2):193–224.
- McFadden D (1974) Conditional logit analysis of qualitative choice behavior. Zarembka P, ed. *Frontiers in Econometrics* (Academic Press, New York), 105–142.
- Meier S, Sprenger C (2009) Discounting and defaulting: Evidence from time preference experiments and administrative credit card data. Working paper, University of California, San Diego, La Jolla.
- Meier S, Sprenger C (2010) Present-biased preferences and credit card borrowing. *Amer. Econom. J.—Appl. Econom.* 2(1):193–210.
- Newey WK, McFadden D (1994) Large sample estimation and hypothesis testing. Engle RF, McFadden DL, eds. *Handbook of Econometrics*, Vol. IV (Elsevier Science, Amsterdam), 2111–2245.
- Nilsson H, Rieskamp J, Wagenmakers E-J (2011) Hierarchical Bayesian parameter estimation for cumulative prospect theory. *J. Math. Psych.* 55(1):84–93.

- O'Donoghue T, Rabin M (1999) Doing it now or later. *Amer. Econom. Rev.* 89(1):103–124.
- Paolacci G, Chandler J, Ipeirotis PG (2010) Running experiments on Amazon Mechanical Turks. *Judgment Decision Making* 5(5): 411–419.
- Phelps ES, Pollak RA (1968) On second-best national saving and game-equilibrium growth. *Rev. Econom. Stud.* 35(2):185–199.
- Prelec D (1998) The probability weighing function. *Econometrica* 66(3):497–527.
- Qiu J, Steiger E-M (2011) Understanding the two components of risk attitudes: An experimental analysis. *Management Sci.* 57(1):193–199.
- Rossi PE, Allenby GM (2003) Bayesian statistics and marketing. *Marketing Sci.* 22(3):304–328.
- Rossi PE, Allenby GM, McCulloch R (2005) *Bayesian Statistics and Marketing* (John Wiley & Sons, New York).
- Sándor Z, Wedel M (2001) Designing conjoint choice experiments using managers' prior beliefs. *J. Marketing Res.* 38(4):430–444.
- Sapienza P, Zingales L, Maestripieri D (2009) Gender differences in financial risk aversion and career choices are affected by testosterone. *Proc. Natl. Acad. Sci. USA* 106(36):15268–15273.
- Sawtooth Software (1996) ACA system for adaptive conjoint analysis. Technical paper, Sawtooth Software, Sequim, WA. Accessed September 7, 2012, <http://www.sawtoothsoftware.com/download/techpap/acatech.pdf>.
- Srinivasan V, Shocker AD (1973) Estimating the weights for multiple attributes in a composite criterion using pairwise judgments. *Psychometrika* 38(4):473–493.
- Stanton SJ, Mullette-Gillman O', McLaurin ED, Kuhn CM, LaBar KS, Platt ML, Huettel SA (2011) Low- and high-testosterone individuals exhibit decreased aversion to economic risk. *Psych. Sci.* 22(4):447–453.
- Starmer C, Sugden R (1991) Does the random-lottery incentive system elicit true preferences? An experimental investigation. *Amer. Econom. Rev.* 81(4):971–978.
- Steinberg DM, Hunter WG (1984) Experimental design: Review and comment. *Technometrics* 26(2):71–97.
- Stott HP (2006) Cumulative prospect theory's functional menagerie. *J. Risk Uncertainty* 32(2):101–130.
- Tanaka T, Camerer CF, Nguyen Q (2010) Risk and time preferences: Linking experimental and household survey data for Vietnam. *Amer. Econom. Rev.* 100(1):557–571.
- Thaler R (1985) Mental accounting and consumer choice. *Marketing Sci.* 4(3):199–214.
- Tom SM, Fox CR, Trepel C, Poldrack RA (2007) The neural basis of loss aversion in decision making under risk. *Science* 315(5811):515–518.
- Toubia O, Evgeniou T, Hauser J (2007a) Optimization-based and machine-learning methods for conjoint analysis: Estimation and question design. Gustafsson A, Herrmann A, Huber F, eds. *Conjoint Measurement: Methods and Applications*, 4th ed. (Springer-Verlag, New York), 231–258.
- Toubia O, Hauser JR, Garcia R (2007b) Probabilistic polyhedral methods for adaptive choice-based conjoint analysis: Theory and application. *Marketing Sci.* 26(5):596–610.
- Toubia O, Hauser JR, Simester DI (2004) Polyhedral methods for adaptive choice-based conjoint analysis. *J. Marketing Res.* 41(1):116–131.
- Toubia O, Simester D, Hauser JR, Dahan E (2003) Fast polyhedral conjoint estimation. *Marketing Sci.* 22(3):274–303.
- Trope Y, Liberman N, Wakslak C (2007) Construal levels and psychological distance: Effects on representation, prediction, evaluation, and behavior. *J. Consumer Psych.* 17(2):83–95.
- Tversky A, Kahneman D (1992) Advances in prospect theory: Cumulative representation of uncertainty. *J. Risk Uncertainty* 5(4):297–323.
- Vohs K, Baumeister RF, Schmeichel BJ, Twenge JM, Nelson NM, Tice DM (2008) Making choices impairs subsequent self-control: A limited-resource account of decision making, self-regulation, and active initiative. *J. Personality Soc. Psych.* 94(5):883–898.
- Wakker P, Deneffe D (1996) Eliciting von Neumann-Morgenstern utilities when probabilities are distorted or unknown. *Management Sci.* 42(8):1131–1150.
- Wang SW, Filiba M, Camerer CF (2010) Dynamically optimized sequential experimentation (DOSE) for estimating economic preference parameters. Working paper, California Institute of Technology, Pasadena.
- Weber EU, Johnson EJ, Milch KF, Brodscholl HCJC, Goldstein DG (2007) Asymmetric discounting in intertemporal choice—A query-theory account. *Psych. Sci.* 18(6):516–523.
- Wu G, Gonzalez R (1996) Curvature of the probability weighting function. *Management Sci.* 42(12):1676–1690.
- Zauberman G (2003) The intertemporal dynamics of consumer lock-in. *J. Consumer Res.* 30(3):405–419.