Dynamically Optimized Sequential Experimentation (DOSE) for Estimating Economic Preference Parameters*

Jonathan Chapman
University of Bologna
jonathan.chapman@unibo.it
www.jnchapman.com

Erik Snowberg
Utah, UBC, CESifo, NBER
snowberg@eccles.utah.edu
eriksnowberg.com

Stephanie W. Wang
University of Pittsburgh
swwang@pitt.edu
www.pitt.edu/∼swwang/

Colin Camerer
Caltech
camerer@hss.caltech.edu
hss.caltech.edu/∼camerer/

September 25, 2025

Abstract

We introduce DOSE—Dynamically Optimized Sequential Experimentation—to elicit preference parameters, and implement it in a representative survey of the U.S. population ($N = 2,000$). DOSE rapidly elicits information about individual preferences by presenting each participant with a customized question sequence. In simulations, DOSE produces parameter estimates that are approximately twice as accurate as those from established elicitation methods. The improvement in accuracy is explained, in large part, by the fact that DOSE accounts for possible participant mistakes. In the field, DOSE is faster to administer, less susceptible to response heuristics, and more stable over time. By reducing measurement error, DOSE identifies a stronger relationship between risk aversion and cognitive ability than other elicitation techniques. These results demonstrate how insights from optimal experimental design can improve measurement, and hence uncover relationships of substantive importance.

JEL Classifications: C81, C9, D03, D81, D9

Keywords: Preference Elicitation, Risk Preferences, Time Preferences, Dynamic Experiments, Cognitive Ability, Preference Stability

*This paper envelops an earlier working paper introducing the DOSE procedure (Wang et al., 2010). We thank Summer Clay, Alison Harris, and Alec Smith for generously sharing their data. Thanks to Roland Benabou, Doug Bernheim, John Beshears, Matthew Chao, Gary Charness, Mark Dean, Andreas Grunewald, Ori Heffetz, Holger Herz, Kate Johnson, Ian Krajbich, Andreas Krause, Sebastian Olschewski, Pietro Ortoleva, Deb Ray, Antonio Rangel, Charlie Sprenger, Stefan Trautmann, Hans-Martin von Gaudecker, Peter Wakker, Nathaniel Wilcox, Leeat Yariv, Horst Zank, and the participants of seminars and conferences for their useful comments and suggestions. Judah Okwuobi and Michelle Filiba provided excellent research assistance. Camerer and Snowberg gratefully acknowledge the financial support of NSF Grant SMA-1329195. Wang thanks the Center for Advanced Study in the Behavioral Sciences.

1 Introduction

Economists have long understood the importance of tailoring information acquisition to aspects of the environment, including a decision-maker's objectives and prior beliefs (Wald, 1947). Yet, experimental methods rarely use this insight when eliciting information about the preferences of individuals. Elicitations are typically chosen using a combination of intuition and precedent, rather than formal criteria. Further, most studies use a one-size-fits all approach, with the same elicitations asked to all individuals within a sample. This approach is convenient for experimenters, and allows straightforward comparisons across studies. However, it may lead to significant inaccuracies in the estimates of each individual's preferences.

In this paper we introduce DOSE (Dynamically Optimized Sequential Experimentation), and investigate, using simulations and field trials, the benefits of this approach for eliciting individual-level preferences. DOSE dynamically selects questions based on a model of preferences, a prior over the parameters of that model, and experimenter-selected criteria. In a parameter recovery exercise, DOSE produces a roughly 50–100% improvement in accurately identifying risk preferences—relative to true parameter values—than established elicitation methods. This improvement in accuracy is due, in particular, to the fact that DOSE selects questions dynamically—incorporating information from previous responses—and that it accounts for the mistakes that individuals may make when answering questions. Consequently, the procedure effectively elicits preferences even if questions are selected with uniform priors or misspecified utility functions. We implement DOSE in a representative, incentivized, online survey, and elicit individual-level measures of risk attitudes, attitudes towards loss, and time discounting. Consistent with the simulation evidence, DOSE appears to produce estimates with lower measurement error than other elicitation procedures. Reducing measurement error uncovers relationships of substantive importance: DOSE identifies greater heterogeneity in economic preferences, stronger patterns of correlations with other individual characteristics, and higher temporal stability of preferences, than other methods.

Eliciting accurate individual-level preference measures is important both for policy and for the study of preferences. In recent decades, economics has moved away from assuming a representative agent, and increasingly recognized the importance of heterogeneity in decision-making (Fagereng et al., 2020; Bilbiie, forthcoming). Economists are studying individual-level behavior in a rapidly expanding range of contexts, fueled by the rise of field experiments, the spread of lab-in-the-field techniques, and growing access to internet survey panels.[1] Parametric measures of preferences, such as those produced by DOSE, can assist policymakers in calibrating macroeconomic models (Fuster et al., 2021), customizing behavioral interventions (Andreoni et al., 2023), discovering undocumented behavioral regularities (Chapman et al., forthcoming), testing theory (Chapman et al., 2023b), and accounting for preference heterogeneity in policy design (Taubinsky and Rees-Jones, 2018; Aguiar et al., forthcoming).

DOSE applies the insights of optimal experimental design to the problem of eliciting economic preferences.[2] As explained in Section 2.2, the procedure starts with a prior probability distribution over parameter values and/or models. An initial question is selected to maximize expected information gain, based on both the prior and an information criterion specified by the experimenter. DOSE then uses the participant's choice to dynamically update beliefs about parameter values and/or models, and selects the next question according to the criterion. The procedure explicitly models the fact that participants may make mistakes, allowing it to recover accurate estimates even if participants choose probabilistically.

[1]Recent studies have sought to identify the distribution of economic preferences in the population (for example, von Gaudecker et al., 2011; Falk et al., 2018), measure correlations between different preferences (Chapman et al., 2023a; Stango and Zinman, 2023), understand the stability of preferences over time (Meier and Sprenger, 2015; Stango and Zinman, 2020; Chapman et al., 2024a), relate economic preferences and psychological traits (Jagelka, 2024), examine relationships between economic preferences and other individual characteristics (Dohmen et al., 2011; Callen et al., 2014), and identify heterogeneity in behavioral decision-making (Allcott et al., 2022).

[2]Optimal experimental design has received greater attention in other fields, particularly in computer science and statistics, than in economics. The idea of dynamic designs can be found as early as Wald (1950); Chaloner and Verdinelli (1995) provide a useful review of applications in statistics. The idea of optimal experimental design appears to originate most clearly in Peirce (1967), who described an "economic" theory of experimentation and applied it to the study of gravity. Optimal designs have been used in many applied fields, including in neurophysiology (Lewi et al., 2009), psychophysics (Kujala and Lukka, 2006; Lesmes et al., 2006), marketing (Toubia et al., 2004; Abernethy et al., 2008), and medicine (Müller et al., 2007). A small number of studies in economics have used optimal experimental design with static, rather than dynamic, experiments (Aigner, 1979; El-Gamal et al., 1993; El-Gamal and Palfrey, 1996; Balietti et al., 2021). See Moffatt (2007) for a discussion of potential applications of optimal design to parameter estimation, including the elicitation of risk preferences.

Two simulation exercises, presented in Section 3, demonstrate that DOSE is more than twice as accurate as standard methods for eliciting risk preferences, and is particularly effective when participants make inconsistent choices. First, we use 140 choices made by each of 120 students (from Sokol-Hessner et al. 2009 and Frydman et al. 2011) to simulate responses to DOSE's adaptive question orderings. We show that twenty DOSE-selected questions capture most of the information from all 140 choices. Importantly, this finding is not sensitive to the prior used by DOSE—using an uninformative prior for question selection produces similar results to using an "optimal" prior that reflects the underlying distribution of parameters. Second, we conduct a parameter recovery exercise with simulated participants, comparing DOSE to two common "one-size-fits-all" elicitation methods—the multiple price list (MPL; see Andersen et al., 2006, for a review), and a risky project measure (Gneezy and Potters, 1997). DOSE produces estimates of risk and loss attitudes that are at least twice as accurate, relative to true parameter values, even if the utility function or error specification used to select questions is misspecified. The improvement in accuracy is particularly large when estimating multiple preference parameters, or when participants are very likely to make mistakes. The latter is demonstrated by showing DOSE may be more than twice as accurate as a simulated partitioning procedure, which chooses questions to successively partition the parameter space without accounting for potential participant mistakes.

We demonstrate the value of DOSE in the field by measuring individual risk and time preferences in a two-wave incentivized, representative study of the U.S. population ($N =$ 2,000), described in Section 4.1. We compare the performance of DOSE to three established elicitation methods—MPLs, a risky project, and qualitative self-assessments. DOSE elicits three preference parameters—risk attitudes, attitudes towards losses, and time discounting—plus a measure of choice consistency in less time than the MPL method takes to elicit a single parameter.[3] Further, participants' behavior does not appear to be distorted by the adaptive or sequential nature of the DOSE procedure.

[3]Pre-programmed question trees used to elicit these parameters are available upon request.

DOSE appears less susceptible to response heuristics—a likely source of non-classical measurement error—than more established elicitation methods. Complex elicitations can lead individuals to rely on heuristics such as *focal value response* (FVR)—choosing salient options such as the top or bottom of lists, the minimum or maximum of continuous response scales (for example, corners of budget lines), or round numbers, such as multiples of 10 (Schwarz and Oyserman, 2001; Barrington-Leigh, 2024). Rates of FVR in our survey are much higher for other methods (32–60%) than for DOSE (less than 3%). Further, consistent with the interpretation that FVR reflects perceived question complexity, the phenomenon is more common among participants with low cognitive ability—potentially confounding estimates of the relationship between cognitive ability and preferences. Moreover, the prevalence of FVR in established methods suggests that our simulations may understate the practical benefits of DOSE.

In Section 5 we illustrate how reducing noise in preference elicitation can uncover findings that would otherwise be obscured by measurement error. A large literature has investigated the distributions of preferences across populations, and the correlations between those preferences and other variables.[4] DOSE facilitates this endeavor as it identifies a much more granular distribution of preference estimates than those obtained by other elicitation methods, and is better able to disentangle different preference parameters, such as separating discounting from utility curvature. Further, we observe significant correlations between a range of individual characteristics and our DOSE-elicited estimates. In contrast, estimated relationships between other elicitations and individual characteristics produces mixed results. For instance, the estimated correlation between cognitive ability and risk aversion ranges from strongly negative (DOSE), to weakly negative (risky project or MPL elicitations), to weakly positive (a qualitative self-assessment from Falk et al. 2018). However, among (only) those participants DOSE identifies as making consistent choices, estimates

[4]For example, several studies have investigated the relationship between risk aversion and cognitive ability (Dohmen et al., 2010), gender Eckel and Grossman (2008), stock market investment Guiso et al. (2018), and education (Hryshko et al., 2011). See Chapman and Fisher (2025) for a broad discussion of the use of preference elicitations.

from the MPL and risky project measures are strongly negatively correlated with cognitive ability. This result suggests that inconsistent choice may explain mixed findings on the relationship between cognitive ability and risk aversion (Dohmen et al., 2018).

Noisy estimates may also explain low temporal stability of preference measures (Meier and Sprenger, 2015; Mata et al., 2018). Our DOSE estimates of participants' risk and time preferences are correlated 0.40–0.47 across a six-month period. In contrast, the correlations for MPL and risky project elicitations range from 0.24–0.33, consistent with these estimates containing more measurement error than those elicited by DOSE. Thus, DOSE answers Meier and Sprenger's (2015, p. 286) challenge to develop, "A more precise experimental technique for eliciting time preferences...to make further study of stability."

We conclude, in Section 6, with a discussion of uses of DOSE. Our results demonstrate that integrating insights from optimal information acquisition can substantially improve preference elicitation, and hence assist researchers in tackling a broad range of research questions. The DOSE procedure can easily be adjusted to investigate other preference types—such as social preferences—or to investigate other aspects of behavior—for instance, learning or beliefs. This, plus the fact that DOSE is fast, accurate, and simple, means that it offers an alternative to experimentally-validated survey modules (Falk et al., 2023) for studying preferences in broad samples. The DOSE procedure thus provides a flexible method that expands the range of settings in which incentivized preference measures can be obtained, and highlights the potential for the use of optimal experimental design in economics.

1.1 Related Literature

Our work contributes to two broad literatures: optimal experimental design, and the measurement of economic preferences and their correlates in broad populations. We review these large literatures only briefly. Relationships between specific prior findings and those in this paper are included when we discuss those specific findings, or in Appendix A.

The DOSE procedure has been implemented in a range of settings based on the initial working paper version of this manuscript (Wang et al., 2010). Most significantly, Chapman et al. (forthcoming) uses DOSE to investigate gain-loss attitudes in the general population.[5] It finds that only around half the U.S. population is loss averse—much lower than in the laboratory samples used in most previous studies. It also demonstrates strong correlations between loss aversion and several behaviors outside of the survey, including gambling, experience of losses, and stock market investment. That paper focuses primarily on a different survey dataset that includes a longer DOSE sequence and other, more traditional, elicitations of loss attitudes. This paper, in contrast, introduces the DOSE method, and uses it to investigate the benefits of the procedure relative to more established methods, in both simulations and online surveys.[6]

Various methods applying optimal experimental design to preference elicitation have been proposed since the introduction of DOSE, but there have been few attempts to evaluate their performance relative to standard, non-adaptive elicitation procedures. Cavagnaro et al. (2010, 2013a,b), and Cavagnaro et al. (2016) apply adaptive design optimization to model discrimination.[7] Toubia et al. (2013) implements a closely related method to study risk and time preferences in a Mechanical Turk sample. Drake et al. (2022) introduce a method to generate adaptive sequences of menus, and prove some convergence properties. These studies focus on explaining the design of each procedure, rather than investigating whether adaptive designs yield significantly better results than non-adaptive designs. This paper, in contrast, evaluates the benefits of DOSE relative to established elicitation methods, and demonstrates that adaptive methods can uncover substantive findings that would otherwise be obscured by imprecise measurement.

Our paper also relates to the broader endeavor to understand the relationships between

[5] Chapman et al. (2018) analyzed the performance of the DOSE procedure and presented results regarding loss tolerance. Based on extensive feedback, we expanded these analyses in separate manuscripts.

[6] Other studies have used DOSE procedure to elicit a variety of preferences in convenience samples, including risk preferences (Clay et al., 2016, 2017), time preferences (Imai and Camerer, 2018; Mitani and Hanaki, 2025), and preferences for autonomy (Freundt et al., 2023).

[7] Their implementations use many more questions than DOSE—80 in Cavagnaro et al. (2016) and 101 in Cavagnaro et al. (2013a)—making it difficult to use in a representative survey.

economic preferences, cognitive ability, and complex preference elicitations.[8] Several studies have concluded that cognitive ability is associated with greater normative rationality (Dohmen et al., 2010; Benjamin et al., 2013), but others have suggested that such results may be confounded by misunderstanding and other measurement issues (Dave et al., 2010; Andersson et al., 2016). Our results complement recent work showing that low cognitive ability is associated with difficulty understanding complex mechanisms for eliciting beliefs (Burdea and Woon, 2022; Burfurd and Wilkening, 2022) This paper demonstrates similar issues in widely-used "simple" elicitation techniques, and shows that they may confound estimated relationships between cognitive ability and other individual characteristics.

2 The DOSE Procedure

This section introduces the DOSE procedure from two perspectives: first from the point of view of participants, then from that of the experimenter. The contrast in perspectives emphasizes that, compared to traditional preference elicitations, DOSE expands experimenters' design choices while streamlining participants' experience. This naturally leads into a discussion of our specific design choices.

2.1 The Participants' Point of View

DOSE is simple from the participants' point of view, as shown in the examples, in Figure 1, taken from our implementation.[9] The participant is given an explanation of the upcoming choices, as in Figure 1a. He or she is then given a series of choices, similar to those in Figure 1b to elicit risk preferences, or Figure 1c to elicit time preferences.

[8]The fact that elicitation methods can affect preference estimates has been well documented (see, for example, Andersen et al., 2006; Crosetto and Filippin, 2016; Pedroni et al., 2017), as have problems with inconsistent choice in MPLs (Filippin and Crosetto, 2016) and other methods (Hey et al., 2009).

[9]This figure includes screenshots from our actual implementation. The instructions page appearing at the start of the DOSE module eliciting time preferences is shown in Appendix Figure F.13. For screenshots of all the questions used in this paper, see Appendix F. Full design documents and screenshots can be found at eriksnowberg.com/wep.

Figure 1: Examples of DOSE from a Participant's Point of View

In the next few questions, you will be asked to choose between two lotteries.

You will start this section with 10,000 points, which you may lose based on the lotteries you choose in this section. That is, some of the lotteries in this section may both **add** to or **subtract** from this initial 10,000 points.

For example, suppose you chose a lottery that had a 50% chance of adding 5,000 points, and a 50% chance of subtracting 5,000 points. In the case of winning, the 5,000 will be added to your additional 10,000. In the case of a loss, the 5,000 will be subtracted from your initial 10,000. Note that you will never have the possibility of losing more than 10,000, so at worst you will end this section with 0 points.

(a) DOSE Instructions

Which of the following options do you prefer?

- A lottery where you can either receive 7,000 points or lose 6,300 points, each with probability 50%;

OR

- Receiving 0 points for certain.

(b) Sample DOSE Choice in Risk-Loss Module

YouGov

Which of the following options do you prefer?

- 10,000 points put in your account 90 days from now (July 16)
- 9,750 points put in your account today

(c) Sample DOSE Choice in Time Module

2.2 The Experimenter's Point of View

DOSE starts with a prior over a set of parameter values, then optimally—according to the experimenter's chosen information criterion—selects questions to pinpoint a participant's preferences. After a participant answers a question, DOSE updates beliefs using Bayes' law, optimally selects the next question, and so on. Consequently, the procedure elicits accurate parameter estimates with only a few questions.

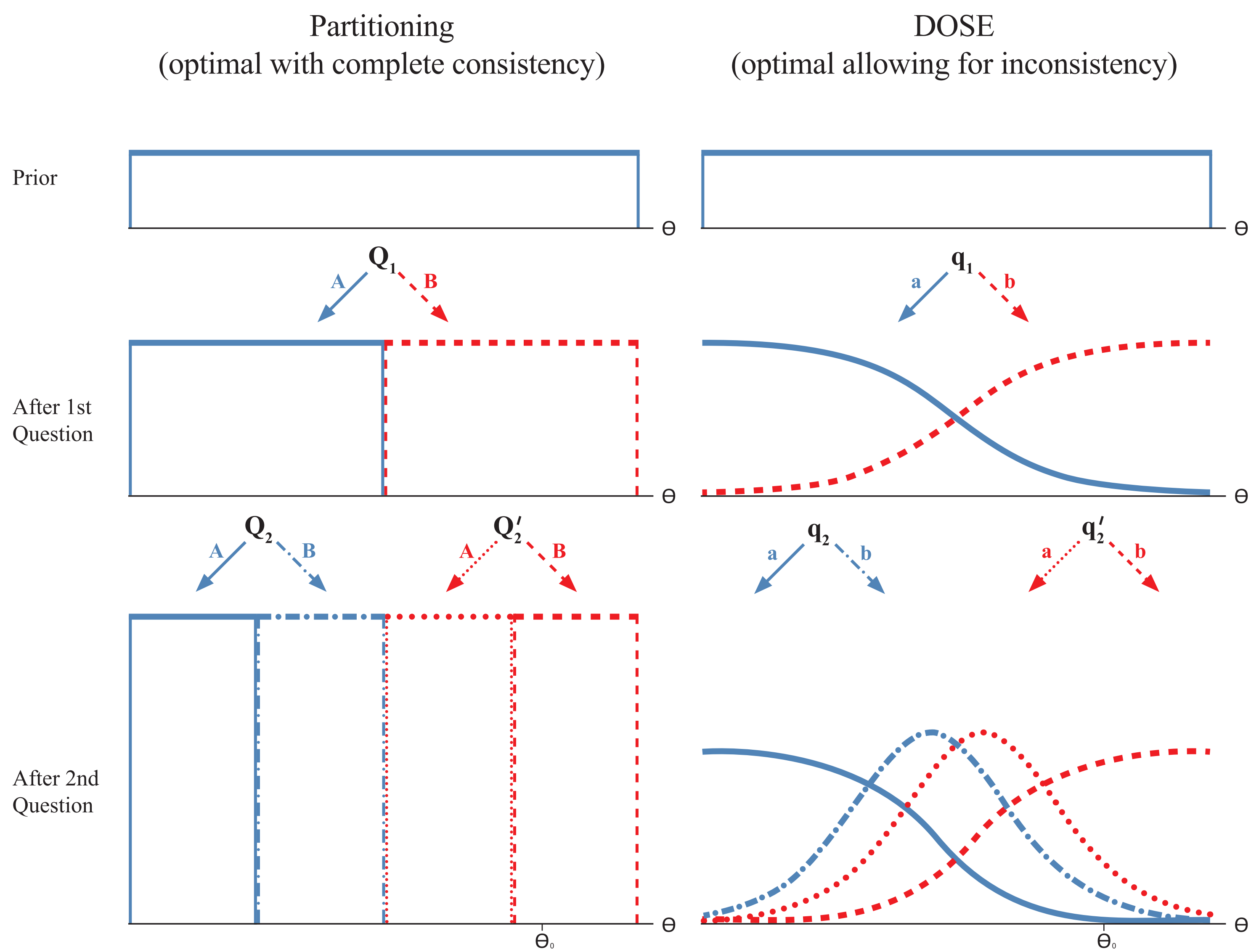
The key difference between DOSE and other common adaptive elicitation methods is that DOSE allows for the possibility that any choice may have been a mistake. We illustrate this, in Figure 2, by comparing DOSE with a simple partitioning method (variously described as bisection, titration, the iterative MPL (see, for example, Andersen et al., 2006), or the staircase method (Falk et al., 2018)).Both methods start with a uniform prior, and offer participants a binary choice. In the first round, all participants face the same question—Q_1 for the partitioning method, or q_1 for DOSE. Beliefs are then updated depending on the participant's answer, and the next question is picked optimally given the new beliefs. In the partition method, parameter values that are inconsistent with a choice, under the assumption of no mistakes, are eliminated from the posterior. DOSE, in contrast, allows for the possibility that any choice may have been a mistake, and hence places a positive probability on all parameter values, regardless of a participant's earlier answers.

DOSE can elicit more accurate parameter estimates than partitioning because it allows for the possibility that participants make mistakes. Consider a participant with true parameter θ_0, displayed at the bottom of Figure 2. In a partitioning method, if the participant incorrectly chooses A in the first question, her estimated parameter value is constrained to be less than the median value—regardless of the number of rounds of questions. Errors early in the procedure thus lead to considerable measurement error. DOSE, on the other hand, places a positive probability on the true parameter value θ_0 even after an initial incorrect choice of a. As a result, with enough correct answers (and perhaps some more mistakes) in future rounds, an accurate parameter estimate will still be obtained.

2.2.1 Our DOSE Procedure for Measuring Risk and Time Preferences

DOSE can be customized for particular research questions. The main objects of choice for a researcher are the parametric specification(s), the prior distribution over parameters or models, the set of choices to present to participants, the mapping from parameters to a distribution over choices—that is, the structure of possible mistakes—and the information criterion used to select questions. Here, we describe the major choices we made when implementing DOSE in our online survey—full details can be found in Appendix B.2. We also highlight aspects of other DOSE implementations, building off our initial working paper, to demonstrate the flexibility of the procedure.

Figure 2: DOSE improves estimate accuracy by allowing for choice inconsistency.



Utility Function and Priors over Parameters. We elicit risk and loss aversion using a Prospect Theory utility function with power utility (Kahneman and Tversky, 1979). This utility function assumes that participants value payments relative to a reference point, which we assume is zero. The standard S-shaped utility function in Prospect Theory implies that participants are risk averse over gains and risk loving over losses. A kink in the utility function at the reference point of zero represents loss aversion. Formally:

$$u(x, \rho_i, \lambda_i) = \begin{cases} x^{\rho_i} & \text{for } x \geq 0 \\ -\lambda_i(-x)^{\rho_i} & \text{for } x < 0, \end{cases} \tag{1}$$

in which λ_i parameterizes loss aversion, ρ_i parameterizes utility curvature, and $x \in \mathbb{R}$ is a monetary outcome relative to the reference point. If $\lambda_i > 1$, then the participant is loss averse. If $\lambda_i < 1$, then the participant is loss tolerant. An individual with $\rho_i < 1$ demonstrates risk aversion over gains and risk love over losses. We measure a participant's risk aversion as $1 - \rho_i$, the *coefficient of relative risk aversion* over gains.

We model time preferences with quasi-hyperbolic discounting and the power utility function in (1). If t is the time from the survey date in months, utility is

$$u(x_t, \rho_i, \beta_i, \delta_i) = \beta_i \delta_i^t x_t^{\rho_i}, \tag{2}$$

in which $\delta_i > 0$ is a discount factor, ρ_i captures the curvature of the utility function from (1), and x_t is a payment at time t. β_i equals 1 for payment in the present ($t = 0$) and—if an individual is present-biased—$\beta_i < 1$ for payment in the future.

We use a joint uniform prior over preference parameters for both question selection and estimation. The support of the prior is chosen based on the individual estimates obtained in Section 3.1. That section also shows this prior leads to near-optimal question selection.[10]

[10] In particular, the prior ranges are $\lambda \in [0, 4.6]$, $\rho \in [0.2, 1.7]$, $\delta \in [0.2, 1]$, $\beta \in [0.4, 1]$, and $\mu \in [0, 8]$—where μ is the choice consistency parameter defined in (3).

DOSE provides additional flexibility because the parametric model used for data selection does not need to be used for ex post data analysis. This feature of DOSE allows researchers to account for information obtained during the experiment (regarding, for instance, the appropriate range of a prior), to use different models for different individuals, or to check the robustness of their results. The simulations in this manuscript show that reanalyzing data ex post allows accurate parameter estimates to be obtained even if the original priors or model are misspecified. Freundt et al. (2023) demonstrate this attribute in practice in using DOSE by identify participants' indifference sets for a lottery. They then show that those sets are almost unchanged under different parametric assumptions about utility functions. Consequently, researchers are not overly restricted by the need to make parametric assumptions when designing an experiment.

Mistakes and Choice Consistency. An important advantage of DOSE is that it models the possibility that participants make mistakes. We model the mapping between utility and choices using the logit function, which has been widely used in both economics and psychophysics due to its connection with the random utility model.[11] For any choice between options o_1 and o_2 with $V(o_1) > V(o_2)$:

$$\text{Prob}[o_1] = \frac{1}{1 + e^{-\mu_i (V(o_1) - V(o_2))}}. \tag{3}$$

The logit function depends both on the utility difference between options o_1 and o_2 and the choice *consistency* parameter $\mu_i \in \mathbb{R}^+$. The probability of making a mistake—that is, not choosing an individual's value-maximizing option—is $1 - \text{Prob}[o_1]$. This is decreasing in the value difference between o_1 and o_2. This decrease is more rapid when μ_i is larger, so higher values of this parameter represent greater consistency in choices.

[11] Specifically, choice probabilities will be logit if the errors in the random utility model have an Extreme Value Type I distribution. See McFadden (2001) for a broader discussion of the history of the logit specification and its properties.

Question Set. DOSE chooses questions from a set determined by the researcher. This set can be very large and, importantly, can include questions that ex ante seem unlikely to be meaningful. In contrast to established methods, which require researchers to fix questions based on their priors, this allows researchers to be agnostic about the type of behaviors they are likely to encounter, and hence be more likely to identify "surprising" outcomes. For example, von Gaudecker et al. 2011 offered participants a fixed set of 56 lotteries, none of which included negative expected values—a reasonable design choice if uniform loss aversion is anticipated, but one that cannot identify the unexpected pattern of loss tolerance we document in Chapman et al. (forthcoming). The DOSE question set we use, in contrast, contains hundreds of possible questions, including choices that can pin down even "extreme" preferences (those that are far from the mean of the prior distribution). In particular, the question set uses binary choices, as we expected this format to be relatively simple for participants (Hey et al., 2009). Further, as we discuss in Section 4.1, these choices were designed to make calculations relatively straightforward. However, the procedure can also be used to select different question types. Imai and Camerer (2018), for example, use DOSE to select Convex Time Budgets (Imai and Camerer, 2018).

Information Criterion. In our implementation, DOSE selects each question to maximize the expected Kullback-Leibler (KL) divergence between the prior and possible posteriors associated with each answer. That is, the question that is picked at each point is the one with the highest expected information gain given current beliefs. The KL criterion has been used widely in the optimal design literature in statistics due to its conceptual simplicity and grounding in information theory (see Ryan et al., 2016, for a discussion and examples). Further, this approach leads to consistent and efficient parameter estimates under weak modeling conditions (Paninski, 2005). However, DOSE can easily be modified to incorporate alternative information criteria—for example, Imai and Camerer (2018) use DOSE with the

EC^2 criterion in a Mechanical Turk sample.[12]

Formally, consider a finite set of possible parameter vectors $\theta_k = (\rho_k, \lambda_k, \delta_k, \mu_k)$ for $k = 1, ..., K$. Each θ_k has an associated probability p_k of being the correct parameters. In the first question, these probabilities are the priors chosen by the experimenter; they are then updated in each round according to the participant's answers. The expected Kullback-Leibler divergence between the prior and the posterior when asking question Q_j is:

$$KL(Q_j) = \sum_{k \le K} \sum_{a \in A} \log \left(\frac{l_k(a; Q_j)}{\sum_{j \in \mathcal{K}} p_j l_j(a; Q_j)} \right) p_k l_k(a; Q_j) \tag{4}$$

where $a \in A$ are the possible answers to the question, and $l_k(a; Q_j)$ is the likelihood of answer a given θ_k—in our implementation this is determined by the logit function in (3). DOSE selects the question that maximizes $KL(Q)$, the participant answers it, model posteriors are updated, the question Q_j that now maximizes $KL(Q)$ is selected, and so on.[13]

Incentive Compatibility. Adaptive designs, including DOSE, are not strictly incentive compatible—in theory, they can be "gamed"—but this is unlikely to be important in practice. Theoretically, a highly sophisticated individual could misrepresent their true preferences in early questions to obtain more generous offers in future questions. For example, participants could misleadingly say they prefer a lottery to a sure amount in the first question in order to increase the magnitude of the sure amounts offered in the future. However, there are many practical and empirical reasons to think that gaming is rare in practice, and will not undermine preference elicitation. Most convincingly, Ray et al. (2012) directly test the importance of gaming in experiments about binary risky choice. In particular, they compare a baseline condition with a condition in which the possibility of gaming was explicitly described. Few

[12]The EC^2 acronym comes from the description "Equivalence Class Edge Cutting" Golovin and Krause (2011), which has been favored in computer science as it has a provable worst-case "cost" bound—that is, the number of choices needed to approximate perfect identification. In addition, the objective function is adaptively submodular, which can economize on real-time computation by re-using previous computations.

[13]We restricted the procedure to only consider questions that had not yet been asked of that participant. In order to improve the estimate of μ, the procedure can be modified to present a question where one option is very likely to represent a mistake (Freundt et al., 2023).

participants attempted to manipulate the algorithm, and those that did so were unable to find an effective method. Moreover, the distributions of best-fit risky choice models did not differ between the control and explicitly-instructed treatment. Further, if participants were attempting to manipulate the procedure, then they should behave differently in earlier rounds (when there is an incentive to act strategically) than later rounds (when there is not)—but they find no evidence that this was the case. See Appendix B.1 for a more detailed discussion.

DOSE can be adapted to be incentive compatible, although our piloting suggests that this would be unlikely to significantly affect results.[14] For example, the DOSE parameter estimates can be used to project each participant's response to an unanswered question (that determines payment).[15] In Wang et al. (2010) we piloted this method in a laboratory experiment. Specifically, half the participants were paid using this incentive-compatible procedure, and half were paid according to a randomly-selected question (as in our online survey). The distribution of estimates was similar across the payment methods, with no statistically significant difference in means. Alternatively, one could randomly select the question chosen for payment from *all* possible questions. If that question is answered in the personalized question sequence, that answer determines payment. If not, the question selected for payment is presented to the participant, and their response determines the payoff (as suggested in Johnson et al., 2021).

Practical Implementation Issues. The DOSE procedure can be adjusted in several ways to assist with implementation in different settings. For example, our survey implementation used a question tree that specified which question should follow a particular answer to a prior question. This tree was generated in advance using simulated choices, as our online

[14] Incentive-compatible designs also have important drawbacks, and care would be warranted in implementing them outside of the laboratory. They are complicated to explain, and there may be a trade-off between the incentive for truthful response and the strength of incentives. The two incentive-compatible payment procedures discussed below, for instance, lower each question's probability of influencing the final payoff.

[15] This procedure was suggested by Ian Krajbich, and implemented in Krajbich et al. (2017).

survey provider could not do the necessary calculations in real time. We also adjusted the procedure to mask the adaptive nature of the question sequence by imposing variation in possible lottery prizes across questions.[16]

3 Performance of DOSE versus Current Methods

In this section, we investigate the performance of DOSE in two sets of simulation exercises, before examining the procedure's effectiveness in a representative survey in the next section. First, we simulate the DOSE question selection procedure using the choices of participants in two laboratory experiments that used questions similar to those in our DOSE implementation. A 20-question DOSE procedure obtains parameter estimates that are, on average, quite close to parameter estimates after 140 randomly-ordered questions. Further, using an uninformative prior for question selection produces similar estimates as a prior that reflects the underlying distribution of parameters. We then use simulated participants to show that, in the presence of mistakes or inconsistency, DOSE recovers parameter estimates that are at least twice as accurate as standard elicitations used to measure risk and loss aversion. Further, misspecification of the utility function used for question selection does not significantly affect the accuracy of the estimates obtained.

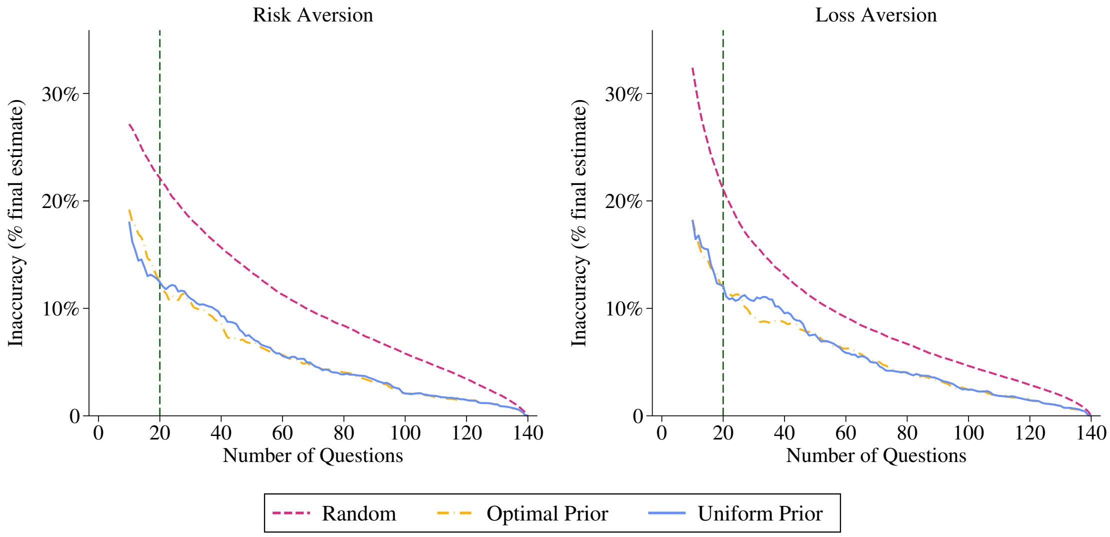
3.1 Simulating DOSE with Laboratory Data

The first simulation exercise illustrates the value of DOSE's personalized question sequence, and investigates the sensitivity of the procedure to the choice of prior used for question selection. We use data from 120 student participants in two prior laboratory experiments.[17]

[16]See Appendix B for more details of the practicalities of survey implementation. Freundt et al. (2023) provides a detailed discussion of adapting DOSE for use on Prolific.

[17]Ninety participants come from Frydman et al. (2011) and thirty from Sokol-Hessner et al. (2009). We attempted to evaluate the performance of DOSE using Maximum Likelihood Estimation (MLE), the method in Sokol-Hessner et al. (2009) and Frydman et al. (2011). However, as reported in Appendix C.3, we were unable to obtain MLE estimates for a large portion of the sample. The MLE estimates that were obtained were less accurate—relative to the estimate after 140 questions—than those obtained from Bayesian estimation.

Figure 3: Optimal question selection rapidly leads to accurate estimates.



Notes: Based on data from Sokol-Hessner et al. (2009) and Frydman et al. (2011). Each line shows the inaccuracy of Bayesian estimates (with uniform initial prior) obtained after each question, starting at question 10, under different orders. "Optimal Prior" and "Uniform Prior" refer to DOSE question selection using corresponding priors. "Random" orders questions randomly. To prevent a particularly extreme random ordering from skewing our results, this line is based on an average over 100 different random orderings.

In each experiment, participants were given a set of 140 binary choices similar to those used in our survey. In our simulation, we order these questions for each participant using DOSE. After DOSE selects a question the answer the participant gave to that question is used to update the probability distribution over parameters. DOSE then selects the next question, and so on. This allows us to compare, question by question, the *inaccuracy*—the absolute distance from the true parameter value as a percentage of the true value—of DOSE's estimates with those elicited by a random question ordering. As we do not have access to true parameter values, we substitute the values one would obtain using all 140 choices.

A 20-question DOSE sequence provides a similar amount of information as about 50 randomly-ordered questions, as shown in Figure 3.[18] The DOSE estimates of both risk and loss aversion are consistently closer to the final parameter estimate, indicating that the procedure provides accurate estimates considerably faster than selecting questions at ran-

[18] For loss aversion, 45 randomly-ordered questions are needed to be as close to the final estimate as 20 DOSE-selected questions. For risk aversion, 55 questions are required. Imai and Camerer (2018) reports similar results relative to fixed or random orders when using DOSE to estimate time preferences.

dom. After 20 questions, the DOSE estimates are almost twice as close to the final estimate as those under a random question ordering (12% vs. 21–22%). The DOSE estimates are also more highly correlated with the final estimates (see Appendix Figure C.1), an important feature when seeking to identify correlations between preferences and other population characteristics.[19]

These simulations also show that using a uniform prior is close to optimal for question selection. We compare the performance of DOSE question selection using a uniform prior to that using an *optimal prior* constructed from the distribution of the estimates after 140 questions. To focus on the question selection impacts of the prior, we estimate the parameter values using a uniform prior in both cases. As shown in Figure 3, the accuracy is similar whether using the optimal or uniform prior. That is, dynamic question selection is effective even if an experimenter has limited knowledge of underlying preferences.

3.2 Parameter Recovery Study

A parameter recovery exercise demonstrates that DOSE produces estimates that are about twice as accurate as traditional risk and loss aversion elicitation methods. We use an entirely simulated dataset that allows us to both know and control the true parameters governing (simulated) participant behavior. Drawing on the insights from the previous simulation, for each simulated participant, we implement a 20-question DOSE sequence, selected using a uniform prior over parameter values. The results indicate that a key feature of DOSE is the fact that it accounts for participant mistakes. Misspecification of the utility function used in question selection does not appear to be a major concern.

We evaluate the relative accuracy of DOSE and three other common risk elicitation procedures using 10,000 simulated participants. Each participant is assigned power utility (as in (1)) with parameter values (ρ_i, λ_i, and μ_i), drawn independently from the posterior

[19]The average benefits we estimate are not limited to the particular distribution of preferences we observe in the laboratory. The DOSE estimates rapidly converge to the final estimate for any given λ and ρ we consider in our simulations, as shown in Appendix C.

Table 1: DOSE produces more accurate estimates than current techniques.

	Average Inaccuracy	Spearman Rank Correlation with True Value
Risk Aversion		
DOSE 10 question	21%	0.66
DOSE 20 question	15%	0.79
Partition	24%	0.62
Multiple Price List	37%	0.45
Risky Project	36%	0.40
Loss Aversion		
DOSE 10 question	21%	0.86
DOSE 20 question	15%	0.91
Partition	36%	0.72
Multiple Price List	36%	0.64

Notes: Inaccuracy is the absolute distance from the true parameter value as a percentage of the true value.

distribution obtained from the 120 laboratory participants in the previous subsection. We run 10- and 20-question DOSE procedures for each simulated participant, with choices generated probabilistically according to (3). Appendix **??** provides full details.

As a benchmark for DOSE, our simulated participants also make choices in three other common risk elicitation methods: investment in a risky project (Gneezy and Potters, 1997), a double MPL (Andersen et al., 2008; Andreoni et al., 2015), and the partition method illustrated earlier in Figure 2.[20] In the risky project, participants are given an endowment that they can partially invest in a lottery which pays a 3x return with 40% probability. In the double MPL, participants complete two MPLs, each offering a choice between a fixed 50/50 lottery and a series of ascending sure amounts. The first MPL—which identifies risk aversion—offers a lottery over gains (\$0 and \$10), while the second—identifying loss

[20] We cannot perform a similar task for the qualitative self-assessments included in our survey as they do not produce parametric estimates. Choices are made using the same logit choice function as in (3). We also simulate a Lottery Menu (Eckel and Grossman, 2002) elicitation procedure, and observe, if anything, slightly worse performance compared to the risky project measure—see Appendix D.1 for details.

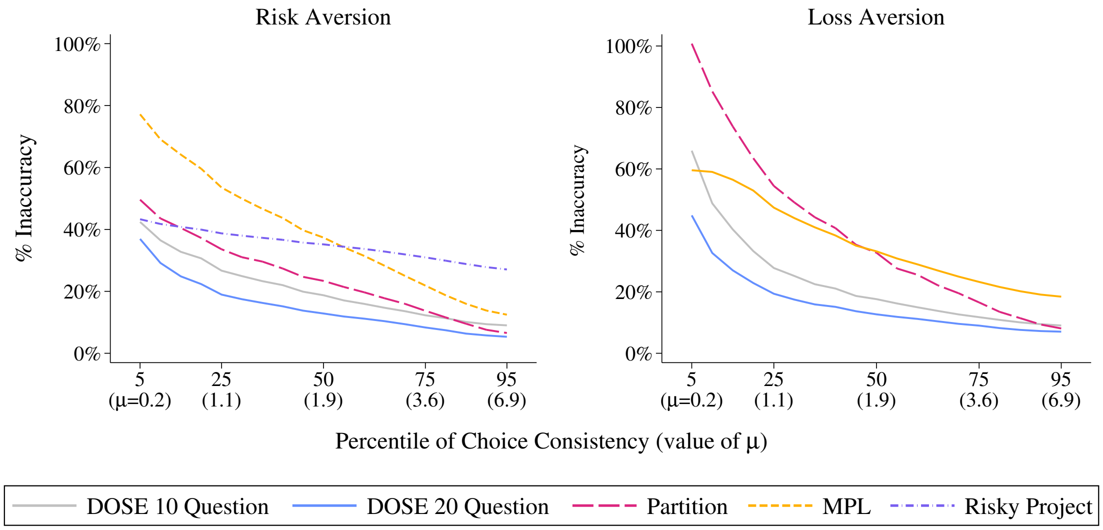
aversion—offers a lottery between a gain of \$10 and a loss of \$10. In both cases, we define a probability distribution over participants' choices by assuming that they consider options until they choose to "switch" from one column of options to the other in the MPL, or select an amount of points to invest in the risky project. This probability distribution is used to calculate the expected inaccuracy of the parameter estimate for each simulated participant. In the partition method, participants face a sequence of twenty binary choices, with each choice selected to bisect the remaining parameter space. The first ten questions estimate risk aversion, and the remaining ten questions estimate loss aversion, accounting for the risk aversion estimate. To obtain parameter estimates, we simulate responses to the question sequence for each participant.

The estimates of risk and loss aversion from DOSE are approximately twice as accurate as those from non-adaptive elicitation procedures, as shown in Table 1. After 20 questions, DOSE obtains estimates that are, on average, within 15% of the true parameter value. The average inaccuracy of the MPL and risky project procedures, in contrast, is at least 35%—much higher than even a 10-question DOSE procedure.

DOSE also performs significantly better than the partition method, demonstrating the importance of accounting for mistakes. The partition method produces more accurate estimates of risk aversion than either the MPL or risky project procedures due to the fact that it also uses a personalized question sequence rather than a one-size-fits-all approach. However, the estimates from partitioning are still over 50% less accurate than those from a DOSE question sequence of equal length. Further, the estimates of loss aversion from the partition method are no more accurate than the double MPL procedure—reflecting the fact that error in the estimates of risk aversion in both procedures are compounded when estimating loss aversion. DOSE thus appears particularly useful when estimating multiple preference parameters.

Importantly, the improvement in accuracy from using DOSE is similar when the utility function used in the question selection procedure is misspecified—see Appendix D.2. We

Figure 4: DOSE is more accurate than other methods at all levels of choice consistency.



Notes: The figure displays estimates obtained from repeating the simulation procedure described in Appendix D 19 times, with each repetition fixing μ at different ventiles of the distribution estimated from asking the full set of 140 questions to the 120 participants discussed in the following subsection.

run DOSE on the same 10,000 simulation subjects—each of whom has CRRA utility—but assume a CARA utility function in the DOSE question selection procedure. Misspecifying the utility function in this way does not lead to a loss of accuracy. Further, even without re-estimating, the Spearman correlation between the estimated CARA parameters and the true (CRRA) parameter values is very high—and notably higher than the correlations for either the MPL or the risky project procedures reported in Table 1. This suggests that assumptions over parametric form are unlikely to be critical if researchers are interested in identifying correlations rather than the level of the risk and loss aversion estimates.

DOSE produces more accurate estimates than the other procedures, regardless of participants' level of choice consistency (μ), as shown in Figure 4. This figure repeats the parameter recovery exercise described above, but assigns all simulated participants the same level of choice consistency, μ. We then vary μ across ventiles of the population distribution. The 20-question DOSE procedure always provides the most accurate estimates, but even a short, 10-question procedure performs better than partitioning, the MPL, or the risky project, except when $\mu \approx 0$, meaning that choice is close to random. DOSE performs particularly well compared to these other procedures in recovering accurate estimates for those below median choice consistency, thus it may be particularly valuable in settings where experimenters have limited experimental control, or among participants most likely to make mistakes.

The high accuracy of the DOSE estimates also leads to higher correlations with the true parameter values than the other procedures (Column 2 of Table 1). Thus, DOSE is less likely to miss associations between economic preferences and other characteristics due to attenuation bias, as we will see in Section 5.3. The correlation between the true risk aversion parameter and the DOSE estimate is 0.79, compared to 0.45 with the MPL estimates and 0.40 for the risky project. The DOSE procedure produces correlations above 0.85 between estimated loss aversion and the true values, even after a 10-question sequence.

Unlike the MPL, DOSE is able to elicit loss aversion estimates even when participants' choices violate First Order Stochastic Dominance (FOSD). As DOSE accounts for the possibility that a participant's choice is a mistake, the procedure can always recover parameter estimates. In the double MPL, on the other hand, participants may erroneously make choices on the second MPL (used to elicit loss aversion) that are first-order stochastically dominated given their choices on the first MPL (used to elicit risk aversion). This prevents estimation of the loss aversion parameter. In our simulation, the MPL could not recover estimates for 11% of participants, increasing to more than 36% of participants in the bottom decile of choice consistency. The next section suggests that these issues are even more common in real-world survey data.

4 Preference Elicitation in a Representative Survey

This section investigates the performance of DOSE relative to some traditional elicitation methods in a multi-wave, online, incentivized, representative survey. We first introduce the survey data, and then demonstrate how the DOSE 20-question module rapidly identifies considerable heterogeneity in preferences. Analysis of survey behavior suggests DOSE was simple to understand, and that the DOSE estimates are not significantly affected by possible survey fatigue or inattention. Moreover, participants appear much less likely to rely on response heuristics in DOSE than in traditional elicitation methods. The next section uses the same data to document how better measurement can uncover substantive findings.

4.1 Data

We implement DOSE in two waves of an incentivized, online survey of the U.S. population ($N = 2,000$).[21] The survey collected a number of behavioral and demographic measures from participants, who were recruited from YouGov's proprietary panel. The two waves asked exactly the same questions of the same participants, about six months apart. Unless otherwise stated, we report results from the first wave.

The behavioral measures in this paper were all incentivized: at the end of the survey, two survey modules were selected at random for payment. All outcomes were expressed in YouGov points, an internal YouGov currency used to pay panel members, which can be converted to U.S. dollars using the approximate rate of $0.001 per point.[22] To enhance the credibility of these incentives, we took advantage of YouGov's relationship with its panel, and restricted the sample to those who had already been paid (in cash or prizes) for their participation in other surveys. The average payment to participants (including the show-up fee) was $9 (9,000 points), which is approximately three times the average for YouGov surveys. The median completion time of the first wave was 40 minutes. We convert points to dollars, using the exchange rate above, when estimating parameter values.

[21] Full details of our specific DOSE implementation are provided in Appendix B. The first wave was conducted between March and April 2015, and the second wave between September and November 2015, both by YouGov. The attrition rate between the two waves of ≈25% is lower than most online surveys, due to YouGov's panel management and the relatively large incentives we offered. To generate a representative sample, YouGov randomly draws people from various Census Bureau products, matches them to members of their panel on observable characteristics, and then provides sample weights to account for differential response rates across sub-populations. We use these weights throughout the paper. This procedure generates better representative samples than traditional probability sampling methods with non-uniform response rates (Pew Research Center, 2016, YouGov is Sample I).

[22] The conversion from points to awards can only be done at specific point values, which leads to a slightly convex payoff schedule. This convexity does not appear to alter participant behavior—see Chapman et al. (forthcoming), especially Online Appendix C.6, for a detailed discussion.

Participants were asked two, 10-question, DOSE modules:

DOSE Risk Preference Module The first DOSE module elicited risk and loss aversion. Participants were given 10,000 points and offered a sequence of ten binary choices between a 50/50 lottery and a sure amount. We selected the lotteries and sure amounts in order to make the choices particularly simple to understand—all lotteries contain 50% probabilities of different payoffs, and there is always a zero involved in the choice, making expected value calculations relatively straightforward. Specifically, two types of lotteries were used. The first had a 50% chance of 0 points, and a 50% chance of winning a (varying) positive amount of points (of up to 10,000). The second had a 50% chance of winning an amount up to 10,000 points, and a 50% chance of a loss of up to 10,000 points. The second type of lottery was always paired with a sure amount of 0 points. The first four questions were restricted to include only the first lottery type, allowing us to obtain a precise estimate of ρ before moving on to estimate λ, and providing a standalone 4-question risk aversion measure.

DOSE Time Preference Module The second DOSE module elicited discount rates and refined estimates of the curvature of the utility function. Participants were offered a sequence of 10 binary choices between a lower amount of points at an earlier date (either the day of the survey, or in the future) or a higher amount at a later date (up to 90 days in the future). The maximum payoff in each question was 10,000 points.

We compare DOSE to methods that vary both in terms of complexity and whether they are incentivized—using the typology of Charness and Viceisza (2016), we include measures that are "incentivized complex" (MPLs), "incentivized simple" (risky project), and "hypothetical simple" (qualitative self-assessments).

Risk Aversion MPLs Two MPLs asked participants to choose between a fixed 50/50 lottery and a series of ascending sure amounts. The row in which the participant first chose the sure amount identified a range of possible certainty equivalents for the lottery—we use the midpoint of this range. There were two MPLs of this type: the first had a 50/50 lottery over 0 and 10,000 points, the second, a 50/50 lottery over 2,000 and 8,000 points.

Risky Project (Gneezy and Potters, 1997) Participants were endowed with 2,000 points that they could keep or invest in a risky project. With 40% probability, the risky project returned 3 times the amount invested, and with 60% probability, the investment was lost. The percentage of the endowment a participant kept measures risk aversion.

Risk Preference Self-Assessment Participants were asked to report their willingness to take risks on a scale of 0 to 10, using the qualitative risk self-assessment from Dohmen et al. (2011), later incorporated in the Preference Survey Module (Falk et al., 2023) and the Global Preference Survey (Falk et al., 2018).[23]

Time Preference MPLs Two MPLs elicited time preferences. The first task elicited the current value of 6,000 points to be received in 45 days. The second elicited the amount of points in 45 days that the participant valued the same as 6,000 points in 90 days. The row in which a participant first chooses to receive money today (rather than in the future) identifies a range of possible discount factors—we use the midpoint of this range. Responses are top-coded to have the same support as the DOSE prior.

Time Preference Self-Assessment Participants were asked to report how likely they are to postpone tasks on a scale of 0 to 10, using a qualitative time preference question taken from the survey in Falk et al. (2018).[24] We code the question so that higher means less likely to postpone, and thus indicates a higher δ or β.

[23]Specifically, participants were asked, "How do you see yourself: are you a person who is generally willing to take risks or do you try to avoid taking risks?" We slightly adapted the wording from the original translation from German reported in Falk et al. (2023) to make the question sound more natural.

[24]Specifically, participants were asked how well the statement, "I tend to postpone things even though it would be better to get them done right away," described them as a person. As it involves only two points in time ("right away" and "postpone"), an increase in "postponement" can be rationalized by a decrease in either δ or β in (2).

Cognitive Ability The survey measured cognitive ability using a set of nine questions. Six questions were from the International Cognitive Ability Resource (ICAR, Condon and Revelle, 2014): three were similar to Raven's Matrices, and the other three involved rotating a shape in space. We also administered the Cognitive Reflection Test (CRT; Frederick, 2005): three arithmetically straightforward questions with an instinctive, but incorrect, answer. Our cognitive ability score was the sum of correct answers to these nine questions (see Chapman et al., forthcoming, for details).

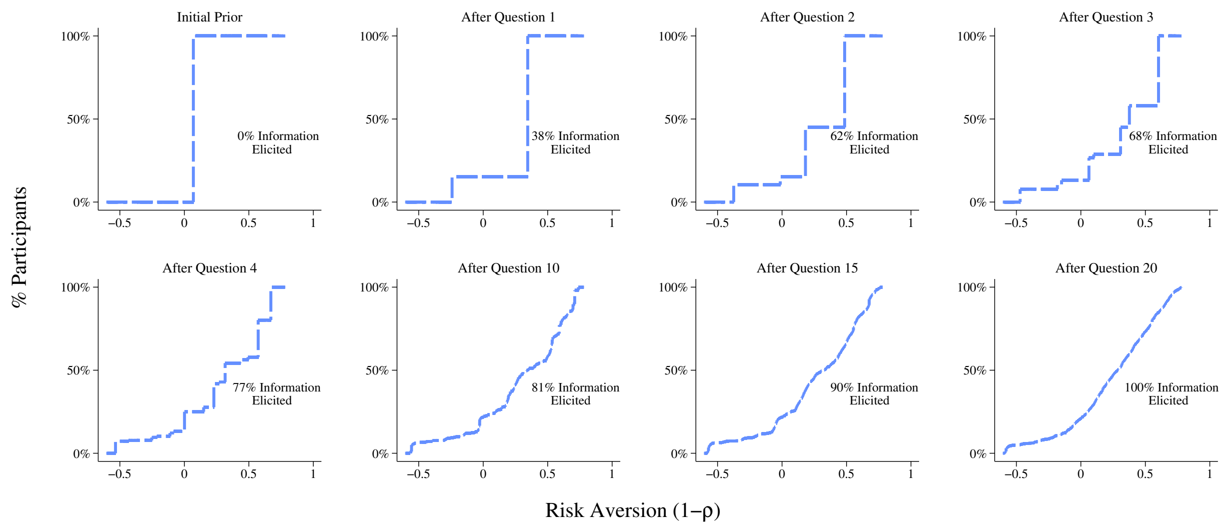
4.2 DOSE in the Field

The DOSE procedure elicited individual-level measures of risk aversion $(1-\rho)$, loss aversion (λ), time discounting (δ), and a measure of choice consistency (μ), in a survey module that took participants an average of 3 minutes and 20 seconds to complete (including instructions).[25] This was approximately one-third faster than the first MPL module on the survey (4 minutes and 25 seconds), which elicited only a measure of discounting.

The DOSE procedure quickly identifies considerable heterogeneity in risk preferences, as shown in Figure 5. The figure plots the cumulative distribution functions of the DOSE estimates of risk aversion $(1-\rho)$ at various points of the DOSE sequence. At the start of the question sequence, shown in the top-left panel, every participant is assigned the same parameter value (specifically, the average of the initial prior). The procedure then splits the sample into two, based on the answer to the first question (second panel), into four after two questions (third panel), and so on. Strikingly, the procedure obtains a great deal of information after just four questions—77% of the distance between the initial prior and the final estimates has been achieved at this point. Further, the DOSE estimates from this 4-question risk preference elicitation—elicited in just 68 seconds, on average—are correlated with cognitive ability (-0.10, with standard error of 0.03), suggesting that even a very short sequence of optimally-selected questions isolates useful information.

[25] We use median response times, as there are a few extremely slow respondents.

Figure 5: DOSE rapidly identifies heterogeneity in risk aversion.



Notes: Each panel displays the cumulative density function of participants' estimated risk aversion $(1 - \rho)$ in the initial prior (top-left) and after further questions from the DOSE question sequence. Initially, every participant is assigned the mean of the initial prior. Subsequently, individual estimates are the mean of the posterior distribution obtained after a given number of questions. "% Information Elicited" reported in the panel is the distance between the CDF in that panel and the CDF after 20 questions, as a % of the distance between the initial prior and the CDF after 20 questions. The distance between CDFs is calculated as the "earth-mover's distance."

Behavior in the survey provides evidence that DOSE works effectively in practice. The fact that DOSE was relatively quick to complete suggests that participants did not struggle to understand the questions offered. Further, an individual DOSE question took 7.2 seconds to complete on average, similar to the time taken to answer the text-based qualitative risk measure (8.8 seconds)—consistent with binary choices being relatively simple to understand. However, one may worry that these fast responses reflect participants failing to pay attention, particularly after receiving a long sequence of questions. We investigate this possibility by examining behavior within the DOSE module, comparing DOSE estimates online with those obtained in the laboratory, and exploring possible order effects.

Participants' behavior during the DOSE modules does not appear to be affected by survey fatigue, either within the DOSE module or across the survey as a whole.[26] At each point of the question sequence, DOSE's prior has an alternative that it "believes" the participant

[26] We present similar exhibits to those reported in this paragraph, using a different dataset, in Chapman et al. (forthcoming, Figure 11 and Appendix C.5).

is more likely to choose. We would expect participant fatigue to cause them to make more random, and thus, "surprising" choices later in the question sequence. In fact, as we see in Appendix Figure E.2, the proportion of surprising choices is lower later in the module. Further, interrupting a 20-question DOSE sequence with a randomly-located page break does not increase the level of surprise—suggesting that the sequential nature of the procedure does not affect participants' choices. We also see that the distribution of DOSE estimates is similar regardless of where DOSE appeared in the survey (Appendix Figure E.1), or how quickly participants completed the survey or the DOSE module (Appendix Figure E.3). Thus, it is unlikely that responses are affected by either fatigue or inattention as a participant moves through the survey.

A comparison of the estimates obtained from different DOSE implementations suggests that participants' behavior is not distorted by the online survey environment, or by learning from repeated exposure to DOSE. DOSE produces estimates in line with previous studies in student populations, whether implemented in the laboratory or in an online survey, as discussed in Appendix A.1. Behavior is similar when participants are presented with multiple DOSE implementations, either across the two waves of our survey (see Appendix A.1), or within a single survey (see Chapman et al., forthcoming). Thus, while we cannot directly examine participants' understanding of the procedure, any learning during the survey module—either due to initial confusion, or recognition that the procedure is adaptive—does not appear to significantly affect estimated parameters.

4.3 Response Heuristics in Preference Elicitations

DOSE appears less susceptible to the use of response heuristics than the other methods in our survey. Complex elicitations can lead to *focal value response* (FVR)—a heuristic whereby participants only consider salient answers such as the bottom, middle, or top of a response scale (Gideon et al., 2017; Barrington-Leigh, 2024). Such a heuristic can create non-classical measurement error, preclude parameter estimates being obtained due to the presence of first-order stochastically dominated choices (see Section 3), and lead to very coarse distributions of estimated preference parameters (see Section 4.2 below). These issues may lead to skewed estimates of correlations between estimated parameters and individual characteristics. Our survey suggests that such heuristics are widespread in established elicitations.

DOSE suffers from much less FVR than established risk elicitation methods suffer, particularly among participants of low cognitive ability, as shown in the left-hand panel of Figure 6. Following Barrington-Leigh (2024), we identify potential FVR by the degree of "heaping" of particularly salient options—the end-point or mid-points of the MPLs, a choice of 0, 5, or 10 in the qualitative self-assessments, or the choice to invest none, half, or all of the endowment in the risky project question. Around one-third of responses in each of the MPLs and in the qualitative risk question (34% and 32% respectively) are at focal values, while in the risky project, 60% of choices were at focal points. FVR is particularly common for participants with low cognitive ability, suggesting that established elicitations place a high cognitive burden on participants.[27]

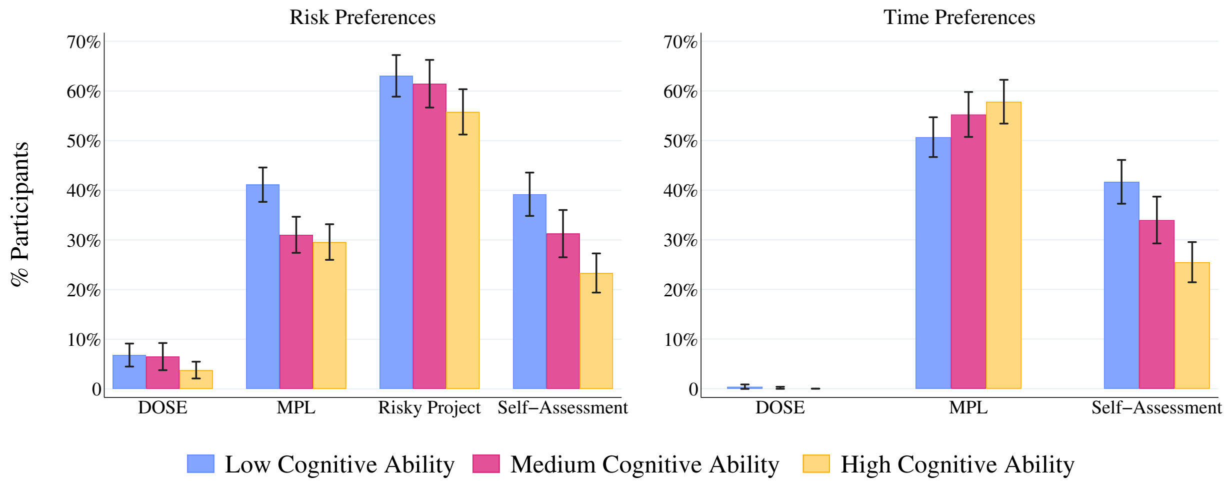
Binary choice tasks, such as those used by DOSE, appear less vulnerable to FVR. Focal values are hard to define in binary choices, but, in principle, participants could focus on only one of the two options—and hence always choose only the first or only the second option (equivalent to consistently choosing either the lottery or the sure amount). Only 6% of participants chose in this way. Thus, the simplicity of binary choices likely reduces non-classical measurement error in preference estimates. However, a nonadaptive design may require an impractically long question sequence (Sokol-Hessner et al. (2009) use more than 140 binary choices). Selecting questions adaptively, using DOSE, mitigates this constraint.

A similar pattern of FVR in the time preference elicitations can be seen in the right-hand panel of Figure 6. In DOSE, very few participants—5 out of 2,000—consistently chose only the first or only the second option in every question.[28] In the time MPL, the rate of FVR

[27] Qualitative self-assessments may be complex due to the need to interpret questions conceptually and project answers onto a numerical scale (Krosnick and Presser, 2010; Chapman et al., 2024b).

[28] In contrast to the risk-loss module, there was no fixed ordering of options in the questions of the time preference module, meaning that we can distinguish choices based on the ordering of the options from the

Figure 6: Choice of focal values is more common for low cognitive ability participants, but is less evident in the DOSE module than in other elicitation methods.



Notes: Figure displays the percentage of participants making focal value choices on each elicitation method. Focal values for each elicitation are: DOSE—always choosing the first or always choosing the second option; MPL—an end-point or mid-point; Risky project—investing none, half, or all of the endowment; Self-reported risk or time—choosing 0, 5, or 10. Bars represent 90% confidence intervals.

was even higher—54% of responses—than in the risk MPLs. However, in this case, the rate of focal value response is highest among those with the greatest cognitive ability—58% in the top tercile, versus 51% in the bottom tercile. This difference may reflect the fact that higher cognitive ability people tend to be more patient, and hence more likely to pick the row of the MPL that indicates no discounting due to a true preference, rather than a heuristic. The pattern for self-assessed time preferences is similar to self-assessed risk attitudes—34% of responses were at focal values. Again, there is heterogeneity by cognitive ability: 41% of those in the bottom tercile chose a focal value, versus 25% of those in the top tercile.

4.4 Measurement Error in Survey Data versus Simulations

FVR was not accounted for in our simulations, implying that the results in Section 3 may underestimate the accuracy gains provided by DOSE. We examine this possibility using data from Chapman et al. (2023a), which employed a survey similar to our main dataset. That survey included two MPLs eliciting risk aversion and an additional MPL eliciting loss

content of those options (that is, early or later payment in this case).

aversion. We simulate these MPLs using the same procedure as in Section 3. The resulting simulation estimates appear to exhibit less noise than the actual survey responses.

We first assess noise using the fact that the share of variance attributable to classical measurement error is, under some assumptions, one minus the correlation between two independent elicitations of the same quantity (Gillen et al., 2019).[29] The correlation between two simulated risk aversion MPLs (0.73) is higher than the correlation in the survey (0.64; s.e. = .02), suggesting there is more noise in the real data. Further, a substantial portion of the correlation in the survey data is explained by participants repeatedly making choices at the end-points of the MPLs—choices which are consistent, but that imply extreme parameter values that are unlikely to be accurate. Excluding such participants, the correlation between the two MPLs falls to 0.43 in the survey data, compared to 0.61 in the simulation.

Measurement error, specifically violations of first-order stochastic dominance (FOSD), is particularly concerning in the double MPL procedure, because it may bias estimates of population parameters, as described in Section 3.2.[30] The survey data indicate that, in practice, this may be a significant issue—37% of responses violate FOSD, almost double the 20% rate in the simulations. Further, violations of FOSD are particularly common among certain subgroups: the double MPL is unable to recover loss aversion estimates for 47% of participants falling in the bottom tercile of cognitive ability, compared to only 30% of participants in the top tercile. As high cognitive ability is associated with more loss aversion (see Section 5.2), this pattern may upwardly bias estimates of loss aversion.

[29] Formally, suppose X^a and X^b are two MPLs eliciting the same underlying preference x^*. Classical measurement error implies that $X^a = X^* + \nu_X^a$ and $X^b = X^* + \nu_X^b$, with ν_X^a, ν_X^b i.i.d. random variables, and $\mathbb{E}\left[\nu_X^a \nu_X^b\right] = 0$. Assuming that $\frac{\mathrm{Var}[\nu_X^a]}{\mathrm{Var}[X^a]} = \frac{\mathrm{Var}\left[\nu_X^b\right]}{\mathrm{Var}[X^b]} \equiv \frac{\mathrm{Var}[\nu_X]}{\mathrm{Var}[X]}$, then

$$\widehat{\mathrm{Corr}}[X^a, X^b] \to_p \mathrm{Corr}[X^a, X^b] = \frac{\sigma_{X^*}^2}{\sigma_{X^*}^2 + \sigma_{\nu_X}^2}.$$

Thus, $1 - \widehat{\mathrm{Corr}}[X^a, X^b]$ estimates the fraction of X^a and X^b explained by classical measurement error.

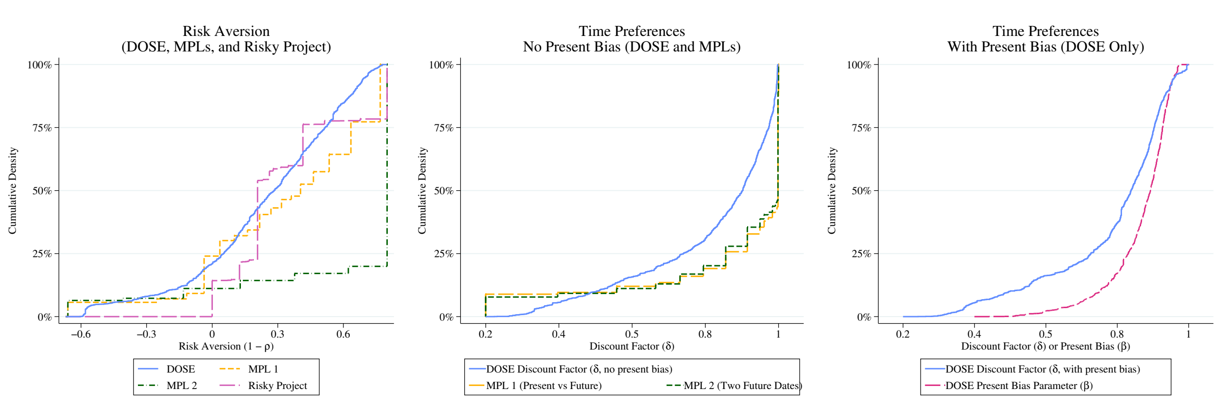
[30] Difficulties with estimating preference parameters due to inconsistent choice are not limited to MPL elicitations. For instance, Booij et al. (2010) drop ~40% of participants due to FOSD violations in a long sequence of binary choices.

The prevalence of FVR and violations of FOSD in the survey data suggest that measurement error poses a serious challenge to the use of established elicitation methods in broad populations. Violations of FOSD can lead to missing data and potential selection bias. Response bunching means that the granularity of estimated preferences may be extremely limited, even when using nearly-continuous scales, such as in the risky project measure. Further, statistical models that treat responses as unobserved latent variables, with errors centered around the latent value, cannot gracefully explain these response patterns—error due to FVR is likely to be systematically correlated with the latent value. Moreover, ex post corrections for classical measurement error, such as the ORIV technique (Gillen et al., 2019), do not apply to non-classical measurement error. The evidence in the next section suggests that these issues may obscure substantive findings.

5 Economic Preferences with Less Noise

Obtaining highly accurate estimates of individual preferences may help researchers uncover relationships that would otherwise be obscured by measurement error. DOSE identifies more heterogeneity in risk and time preferences than other elicitation measures. The choice data generated by the procedure can be used to test robustness to alternative parametric assumptions, or simply examine preferences non-parametrically. Our DOSE-elicited parameter estimates are robustly correlated with a number of individual characteristics, including cognitive ability. This is in contrast to other incentivized elicitations of risk aversion, which show patterns of much weaker correlations. This difference is explained, in part, by the fact that DOSE accounts for inconsistent choice—we observe similar patterns of correlations for all three incentivized elicitation methods when excluding participants who DOSE estimates to have below-median choice consistency (μ). Finally, DOSE estimates of economic preferences are more stable over time than estimates from traditional incentivized elicitations.

Figure 7: DOSE identified more variation in risk and time preferences than other elicitation techniques.



Notes: The left-hand panel displays the cumulative density functions for estimated risk aversion (1-ρ) from the 20-question DOSE procedure, from two MPL elicitations offering a choice between a lottery and a sure amount, and from a risky project elicitation. The middle panel displays cumulative density functions for estimated discount factors (δ) from the DOSE procedure with β set equal to 1, and from the two MPL elicitations. The first MPL ("Present vs Future") offered a choice between payment today or in 45 days. The second MPL ("Two Future Dates") offered a choice between payment in 45 days or in 90 days. The estimated parameters for all elicitations are restricted to the range of the relevant DOSE prior. The right-hand panel displays the cumulative density function for estimates of (δ) and (β) from DOSE when allowing for present bias. It was not possible to obtain meaningful estimates of β from the MPLs due to the similarity of responses across the two elicitations.

5.1 Identifying Heterogeneity in Economic Preferences

The DOSE procedure identifies greater heterogeneity in both risk aversion and time discounting than other elicitation methods, as shown in Figure 7. The left-hand panel shows that the distribution of estimated risk aversion from DOSE is less lumpy than that elicited by either MPLs or the risky project measure. The middle panel compares the distribution of the exponential discount factor, fixing $\beta = 1$ for all participants, obtained from DOSE and from two MPLs. Conceptually, DOSE improves on the MPL measures because it accounts for variation in utility curvature. Moreover, in the MPLs, a majority of participants made choices indicating no discounting ($\delta = 1$). Further, as discussed in Section 4.3, such choices may reflect heuristics, rather than an individual's true preferences.

DOSE is particularly effective at uncovering heterogeneity in preferences from multi-parameter models, as we demonstrate in the right-hand panel of Figure 7. Because the procedure offers participants many different choices, the data it generates can be used to investigate alternative functional forms. In this figure, we present the results from re-estimating the discount function in (2), with β no longer fixed. Again, we see considerable heterogeneity. However, we could not estimate a β parameter from the MPL module, due to the similarity of response patterns across the two MPL elicitations. This finding suggests that results in previous studies indicating no present bias when using monetary rewards could, at least in part, be an artifact of the elicitation method—in particular, the tendency for participants to cluster at particular values. It has been suggested that apparent present bias could reflect payment risks rather than hyperbolic discounting (Andreoni and Sprenger, 2012). However, in our results, the existence of present bias is determined by the elicitation type, as any perceived payment risk is the same across all our elicitations.[31]

Our findings regarding loss aversion, detailed in Chapman et al. (forthcoming), illustrate how DOSE's flexibility can aid new discoveries and enable rigorous robustness checks. The main finding of that paper is that only around 50% of the U.S. population is loss averse—a pattern we did not anticipate when launching our initial survey. However, because DOSE's prior allowed for values of $\lambda < 1$, and the set of possible questions included many that allowed for $\lambda < 1$—specifically choices between \$0 for sure and lotteries with negative expected value—we were also able to identify "loss tolerant" individuals. Further, the significant volume of choice data collected by DOSE allowed us to verify that this finding was not an artifact of our assumptions.

5.2 Correlates of Economic Preferences

We observe robust correlations between all of our DOSE-elicited parameter estimates—both risk and time preferences, as well as choice consistency—and individual characteristics and behaviors. More cognitively able individuals tend to be less risk averse, more loss averse,

[31] Andreoni and Sprenger (2012) find limited evidence of present bias when using either a double MPL method or Convex Time Budgets (CTBs). Participants also tend to choose potentially focal values in CTBs—in their experiment, 70% of all choices were focal values (corner choices), and 37% of participants chose a focal value in every choice. See Chakraborty et al. (2017) and Imai and Camerer (2018) for detailed discussions of corner choices and other evidence of inconsistent behavior within CTBs.

more patient, and make more consistent choices than those with lower cognitive ability. Correlations with other characteristics largely match the patterns found in prior studies, and are robust to allowing for potential misspecification of the utility or error function.

Table 2 reports correlations between our four DOSE-elicited parameter estimates and a range of individual characteristics. For each preference parameter, we first report univariate correlations with each characteristic, and then report multivariate regressions that include all characteristics simultaneously. The results for risk aversion and loss aversion reproduce findings reported in Chapman et al. forthcoming, while those for patience and choice consistency are novel to this paper.

The correlations in Table 2 are largely consistent with extant literature.[32] In our data, more educated, higher income, and more cognitively able individuals tend to be less risk averse, more loss averse, more patient, and make more consistent choices. Cognitive ability is the strongest predictor of each parameter, with correlations that are robust to inclusion of controls. Notably, in fact, cognitive ability appears to explain a large part of the relationship between education and economic preferences.[33]

Moreover, these results are robust to checks for misspecification of either the utility or error function used to produce DOSE estimates, as detailed in Appendix E.3. The DOSE estimates are underpinned by a clear pattern of choices: higher cognitive ability participants are more likely to accept lotteries offering only potential gains, less likely to accept lotteries with potential losses, and more likely to accept later-dated payments.

[32] See Appendix A for a more detailed discussion of prior studies examining the correlates of economic preferences.

[33] Appendix E.4 shows that the results in Table 2 are similar when allowing for a non-monotonic relationship between economic preferences and cognitive ability, including using categorical dummies for control variables, and restricting the sample to those above median response time (on DOSE or the survey as a whole) or those above median choice consistency. Appendix Table E.5 demonstrates that it is the inclusion of cognitive ability that attenuates the coefficient related to education.

Table 2: Cognitive ability is the strongest correlate of economic preferences.

	Risk Aversion $(1-\rho)$		Loss Aversion (λ)		Patience (δ)		Choice Consistency (μ)	
	Univariate Correlation	Multivariate Regression	Univariate Correlation	Multivariate Regression	Univariate Correlation	Multivariate Regression	Univariate Correlation	Multivariate Regression
Cognitive Ability	-0.21***	-0.18***	0.21***	0.16***	0.18***	0.17***	0.16***	0.15***
	(0.028)	(0.030)	(0.030)	(0.033)	(0.029)	(0.028)	(0.026)	(0.029)
Education	-0.10***	-0.02	0.13***	0.05	0.17***	0.10***	0.11***	0.06*
	(0.033)	(0.034)	(0.032)	(0.034)	(0.037)	(0.038)	(0.031)	(0.032)
Income	-0.11***	-0.07	0.13***	0.08**	0.10***	0.00	0.08**	-0.01
	(0.035)	(0.042)	(0.030)	(0.040)	(0.035)	(0.039)	(0.035)	(0.041)
Male	-0.10***	-0.06*	0.07**	0.03	-0.02	-0.05*	0.00	-0.03
	(0.032)	(0.031)	(0.033)	(0.032)	(0.035)	(0.033)	(0.033)	(0.033)
Age	0.01	-0.00	-0.10***	-0.11***	0.18***	0.18***	0.06*	0.05
	(0.032)	(0.032)	(0.033)	(0.033)	(0.036)	(0.036)	(0.036)	(0.035)
Married	-0.01	0.02	0.04	0.02	0.09***	0.02	0.11***	0.09***
	(0.032)	(0.037)	(0.034)	(0.038)	(0.035)	(0.038)	(0.033)	(0.036)
N	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000

Notes: ***, **, * denote statistical significance at the 1%, 5%, and 10% level. Robust standard errors, in parentheses, come from a standardized regression. The first, third, fifth, and seventh columns report univariate correlations. The second, fourth, sixth, and eighth columns report the coefficient from a multivariate regression.

These patterns mean that our estimates of participants' preferences—and the estimated correlations with other individual characteristics—are very similar when re-estimating with alternative parametric assumptions. In particular, the correlation between cognitive ability and discounting is almost identical when allowing for present bias. The correlations between cognitive ability and both risk and loss aversion are very similar when re-estimating the parameter estimates assuming Constant Average Risk Aversion (CARA) utility, allowing for differential curvature over gains or losses, modeling choice consistency using a probit function, or when using the Random Parameter Model suggested by Apesteguia and Ballester (2018).[34]

5.3 Choice Consistency and Estimate Accuracy

Inconsistent choice can explain much of the mixed evidence regarding the relationship between risk aversion and cognitive ability both within our survey and in the literature (Dohmen et al., 2018). As demonstrated by the simulations in Section 3, MPLs and other techniques measure risk aversion with considerable error when participants make inconsistent choices. Error attenuates, and potentially biases, any estimated relationship between these measures and other factors. In this subsection, we show that inconsistent choice is related to attenuation in our survey. In particular, DOSE identifies stronger correlations between individual characteristics and economic preferences than established risk aversion measures. However, for incentivized elicitation methods, the pattern of correlations are more similar when we focus only on participants that make (more) consistent choices.

The estimated correlation between risk aversion and cognitive ability varies considerably across elicitation methods, as shown in the right panel of Figure 8. The correlation with the estimate from the full 20-question DOSE sequence is -0.22 (s.e.=.03). In contrast, the correlation with the MPL-based risk aversion measure is just -0.01 (.03), with the risky

[34] In fact, we observe a similar pattern of correlations between cognitive ability and non-parametric preference measures derived from the DOSE choice data. Cognitive ability is correlated 0.13 (s.e. = .03) with the share of lotteries over only gains accepted by a participant, -0.21 (.03) with the share of mixed lotteries accepted, and 0.22 (.02) with the share of later-dated payments accepted. Estimating risk aversion using the 4-question DOSE sequence we find a correlation of -0.10 (.03) with cognitive ability, using the 10-question risk-loss module the correlation is -0.12 (.03).

project measure it is -0.10 (.03), and with the qualitative self-assessment it is 0.01 (.03). Thus, depending on the measure, we find a correlation ranging from strongly negative (and statistically significant) to slightly positive (and insignificant). These patterns are consistent with the simulation findings that DOSE obtains more accurate estimates than MPLs—and hence correlations with DOSE estimates suffer less from attenuation bias.

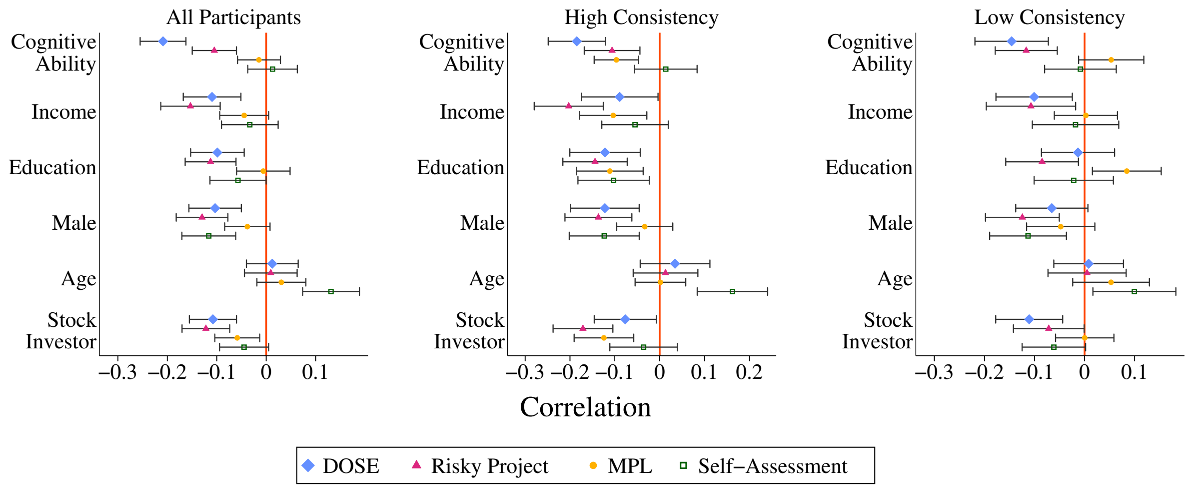
Further, the pattern of correlations for all three incentivized elicitation methods is similar if we focus on participants with above-median choice consistency (μ)—see the middle panel of Figure 8.[35] A number of relationships between the MPL measure and other variables—including cognitive ability—are statistically significant, despite larger standard errors due to the smaller sample. The DOSE estimates, in contrast, exhibit similar correlations when limiting the sample to participants with below-median choice consistency—except education, right-hand panel—while correlations with MPL estimates are near zero. Correlations with the self-assessment question, however, have key differences in all three panels, implying that this measure is capturing different information (Chapman et al., 2024b).

We also observe different correlations between cognitive ability and time preferences depending on the measure (see Appendix Table E.3 for details). In particular, cognitive ability is positively correlated with both DOSE (0.18; s.e.=.03) and MPL measures (0.19; s.e.=.03) of discounting (δ), and also with the DOSE measure of present bias (0.17; s.e.=.03—see Appendix Table E.2). However, cognitive ability is negatively correlated with the qualitative self-assessment of time preferences (-0.10; s.e.=.03), implying that individuals with higher cognitive ability place a lower value on the future, in stark contrast with the results from the DOSE and MPL-based measures of time preferences.

The patterns in Figure 8 demonstrate that the DOSE choice consistency parameter can

[35]For this analysis, we identify choice consistency using only the time preference DOSE module, which seems to produce more informative estimates of μ, because the time questions included responses which are clear "mistakes" under most parameter values—for instance, accepting substantially fewer points today than tomorrow. The results are similar when splitting the sample using the μ from the entire question sequence—see Appendix Figure E.7. For comparability to DOSE, Figure 8 uses risk aversion parameters derived from each incentivized elicitation, as in Figure 7. The results are very similar when using non-parametric measures—certainty equivalents from the MPLs, or the amount invested for risky project—see Appendix Figure E.8.

Figure 8: DOSE measure of risk aversion is more highly correlated with individual characteristics before accounting for choice consistency.



Notes: Figure displays correlations between each measure of risk aversion and individual characteristics. The left-hand panel includes all participants, the middle contains those with above median choice consistency, and the right-hand panel contains those with below median choice consistency. The survey contained two MPL measures of risk preference—we estimate correlation coefficients using stacked regressions and clustering standard errors by participant. Bars represent 90% confidence intervals.

assist researchers in addressing survey noise by identifying individuals who, on that survey, are likely making more mistakes—even outside of the DOSE module. Such information is difficult to obtain from other readily available measures. The correlations in Figure 8 cannot, for example, be recovered by truncating the sample based on response time rather than consistency (Rubinstein, 2016) (see Appendix E.5). In fact, the consistency measure distinguishes whether fast responses reflect inattention: restricting the sample to high-consistency participants recovers correlations even among participants with fast response times. This pattern is particularly striking given that choice consistency was not the primary focus of our DOSE design, and minor design tweaks could produce an even more useful estimate of μ (Freundt et al., 2023). As such, the DOSE consistency parameter may provide an alternative to directly eliciting perceived question complexity (Agranov et al., 2025).

Moreover, we cannot recover similar correlations between cognitive ability and risk aversion by applying techniques to account for classical measurement error. The fact that our survey contains duplicate MPL elicitations allows us to use the ORIV technique (Gillen et al., 2019). This technique produces estimates that account for measurement error that is orthogonal between the two MPL elicitations. The estimated ORIV correlation for the MPL risk aversion measure is -0.05 (.03), smaller than the correlation obtained using the risk parameter from a 4-question DOSE sequence. These results suggest that other risk aversion elicitations contain sources of non-classical measurement error—such as the use of the FVR heuristic documented above—that DOSE avoids.

5.4 Within-person Stability of Economic Preferences

The within-person stability of the DOSE estimates of risk and time preferences is higher than other incentivized behavioral elicitations, consistent with the fact that DOSE reduces measurement error in parameter estimates. The correlation of DOSE estimates within-person across survey waves was 0.40 (s.e. = .04) for loss aversion (λ), 0.44 (.04) for risk aversion (ρ), and 0.47 (.05) for discounting (δ), as reported in Chapman et al. (forthcoming). The correlation of the ρ parameter estimated with choices from just the first four questions, 0.44, was similar to that after 20 questions. This reflects the fact that much of the information about that parameter was elicited in the first four questions, as shown in Figure 5.

Established incentivized elicitation methods produce estimates with significantly lower over-time stability than DOSE. The risky project measure had inter-temporal stability of 0.33 (.04), and the stability of our MPL measures was even lower—the inter-temporal correlations of choices across two risk MPLs were 0.29 (.04) and 0.26 (.04), and in two time preference MPLs were 0.28 (.06) and 0.20 (.06). We also find similar magnitudes of correlations for a broader range of risk elicitations in Chapman et al. (2024a). These findings are consistent with our earlier results indicating higher measurement error in MPLs.

The over-time correlations of the DOSE estimates are also higher than those found with established incentivized elicitation methods in previous studies (see discussion in Appendix A). Meier and Sprenger (2015) report correlations over one year of 0.36 for present bias and 0.25 for discounting parameters among 250 low- to middle-income Americans. Stango and Zinman (2020) study a more representative sample using hypothetical measures, and report rank-stability of patience (0.29) and risk aversion (0.38). Notably, the correlations of DOSE estimates are on the high-end of prior results even when compared with studies in more controlled laboratory environments: Frey et al. (2017) report an average correlation of 0.46 across eight behavioral risk measures (including an MPL) among a sample of 107 adults, while Gillen et al. (2019) find an inter-temporal correlation of 0.32 for each of two risk MPLs, and 0.36 and 0.47 for two risky project questions in a sample of 786 Caltech students.[36] This comparison suggests that DOSE obtains similarly low levels of measurement error in an online survey as existing methods do in more controlled environments.

Our findings suggest that limitations of established elicitation methods may, at least in part, explain why incentivized preference measures tend to have lower over-time correlations than self-reported measures. The self-reported measures in our survey have higher intertemporal stability than the incentivized elicitations—the correlation of the self-reported risk measure is 0.58 (.04), while that of self-assessed time preferences is 0.69 (.03). This pattern is consistent with previous findings in the literature. Mata et al. (2018) carry out a meta-analysis, and find that self-reported risk measures have over-time correlations of around 0.5 over periods of up to ten years, whereas for lottery-based measures (such as those we examine) correlations are on average around 0.2.[37] Our findings suggest that measurement error in established elicitation methods may lead to the underestimation of the over-time stability of economic preferences.

[36] Three papers have studied stability using incentivized elicitations in small laboratory samples over lengthy time periods. Levin et al. (2007) ($N = 62$) report over-time correlations of 0.29 for risk-taking over gains, and 0.20 for differential risk-taking between the gain and loss domain. Lönnqvist et al. (2015) ($N = 43$) report a within-subject correlation of 0.21 for an MPL risk measure across a year. Kirby (2009) ($N = 46$) reports correlations of 0.63–0.71 for a discounting measure.

[37] Preuss (2021) reports a similar magnitude of over-time correlations for self-reported patience (0.50) and impulsiveness (0.44) over a five-year period in the German SOEP.

6 Discussion

This paper moves experimental elicitation of preferences closer to theories of optimal information acquisition with the introduction of DOSE—Dynamically Optimized Sequential Experimentation. Our main results are summarized in Table 3. First, DOSE elicits multiple fine-grained estimates of preference parameters in the time it takes established elicitations to elicit one much coarser estimate. Second, in contrast with other methods, our DOSE implementation does not suffer from significant rates of focal value response, suggesting that the procedure is easy to understand, so participants do not use heuristics. Third, the DOSE parameter estimates are more highly correlated with cognitive ability and more stable over time than standard elicitations, in line with our simulation evidence that the latter suffer from greater measurement error. Fourth, among participants with high choice consistency, as identified by DOSE, there is a strong negative correlation between higher cognitive ability and other incentivized measures of risk aversion. Together, these results suggest that measurement issues may explain both low estimates of inter-temporal stability of economic preferences, and mixed results about the relationship between cognitive ability and risk aversion in the extant literature.

Our results also shed light on the most important aspects of optimal information acquisition in preference elicitation. When eliciting risk and time preferences, DOSE's improvement in accuracy over established methods is largely driven by selecting questions dynamically—incorporating information from previous responses—and accounting for the mistakes that individuals may make when answering questions. These advantages are not reduced by using a naive prior or misspecified utility functions. Further, DOSE facilitates the use of simple binary choices to elicit preferences, which largely eliminates the use of the FVR response heuristic.

Table 3: Comparison of DOSE with Other Incentivized Elicitation Methods

	DOSE	MPL	Risky Project	Self-Assessments
Correlations w/ Demographics				
Risk Aversion	✓	≈ 0	✓	≈ 0
Loss Aversion	✓	n.a.†	n.a.	n.a.
Patience / Time Preference	✓	✓	n.a.	✓
Correlations w/Cognitive Ability				
Risk Aversion	-0.21	≈ 0	-0.07	≈ 0
Loss Aversion	0.40	n.a.†	n.a.	n.a.
Patience / Time Preference	0.18	0.18	n.a.	-0.10
Parameter Recovery Analysis				
Inaccuracy: Risk Aversion	15%	37%	36%	n.a.
Correlation: Risk Aversion	0.79	0.45	0.40	n.a.
Inaccuracy: Loss Aversion	15%	36%	n.a.	n.a.
Correlation: Loss Aversion	0.91	0.64	n.a.	n.a.
Survey Behavior				
Speed	115 secs	259 secs	33 secs	12 secs
Stability: Risk Aversion	0.44	0.28	0.33	0.58
Stability: Loss Aversion	0.40	n.a.†	n.a.	n.a.
Stability: Patience	0.47	0.24	n.a.	0.69
Evidence of Heuristics				
Choice of Focal Values	3%	34%	60%	33%
Correlation w/Cognitive Ability	-0.07	-0.11	-0.07	-0.17

Notes: Correlations with cognitive ability and with demographics are estimated using our survey data, with ✓ indicating that a pattern of statistically significant correlations was identified (Section 5.2). "Stability" is the correlation across survey waves, conducted six months apart (Section 5.4). Parameter Recovery Analysis (Section 3.2 and Appendix D) tests how well a procedure recovers known parameters from simulated data: "Inaccuracy" is the average absolute percentage difference between the estimated and true parameter values, and "Correlation" is the correlation between the estimated and true parameter values. While most entries of n.a. are because the elicitation type is not designed to measure the quantity of interest, † indicates that estimates for loss aversion could not be recovered because of violations of FOSD, see Section 3.2. Speed represents the survey time for a single survey module (for DOSE, this measures the time taken to elicit risk and loss aversion; for the MPL, this measures the time taken to complete the time preference MPL module, including reading instructions), or, in the case of self-reports, a single question. Evidence of Heuristics refers to the possible use of Focal Value Response, discussed in Section 4.3 (for the MPL and risky project, this is measured using the risk aversion elicitation, and for DOSE and self-assessments, it is measured using both risk and time preference elicitations).

It is straightforward to adapt the DOSE procedure to tackle many other research questions. The procedure could be used to elicit willingness-to-pay, or alternative sets of behavioral parameters, such as probability weighting functions or social preferences (Camerer et al., 2020). Freundt et al. (2023), for example, use DOSE as part of a two-step procedure to estimate individual preferences for choice autonomy, and use the DOSE choice consistency parameter to identify highly consistent participants. Alternatively, slight modifications of the DOSE implementation used here could investigate "consistent inconsistencies" in the choices of participants, such as the use of focal value response, and hence account for such choices in parameter estimates.[38] Future researchers interested in incentive (in)compatibility in dynamic processes could adapt the DOSE procedure to incorporate the possibility of strategic manipulation as a separate theory of behavior, drawing on the ideas discussed in Section 2.2 and Appendix B.1 (see Echenique and Prasad 2019 for a theoretical discussion of incentive-compatible learning algorithms). More generally, adaptive methods such as DOSE can expedite the evolution of rapid elicitation procedures that capture multiple preferences simultaneously, akin to the transition in psychology from long-form (50-100 items) to shorter-form scales (5-10 items) that allow many characteristics to be measured simultaneously.

Critically, the accuracy, speed, and simplicity of DOSE potentially expands the range of research settings in which incentivized preference elicitation is viable. In particular, economists have often shied away from using incentivized measures in large samples because of high measurement error and the prohibitive cost of implementing multiple elicitations (Schildberg-Hörisch, 2018). DOSE may be particularly valuable in field experiments, where it is difficult to provide participants with detailed instructions. Similarly, it may also be easier to implement than more complex or time-consuming designs when conducting experiments with low-literacy participants, children, patients with medical disorders, or even animals. DOSE performs better than other elicitations when studying low-cognitive-ability,

[38]To identify the importance of focal value response, researchers could compare parameter estimates when using question sets with and without round numbers, or test whether starting a question sequence with a round number leads to anchoring effects.

low-education, and low-income participants, suggesting that it could be particularly useful in developing countries, where current elicitation methods can be undermined by inconsistent choice (Jacobson and Petrie, 2009; Charness and Viceisza, 2016), and new techniques are especially needed (Berry et al., 2020). Indeed, our finding of widespread loss tolerance suggests that the failure to investigate the preferences of such groups may lead important behavioral regularities to be overlooked (see Chapman et al., forthcoming).

The ability to accurately estimate individual-level preference parameters in the field can contribute both to the study of preferences and the design of behavioral policy. More accurate measurement of preferences opens the path for deeper understanding of how preferences are shaped, and how they influence individual behaviors. Individual-level preference estimates could be incorporated into personalized, targeted—and hence more effective—policy interventions, such as customized incentivized contracts Andreoni et al. (2023) or adaptive treatment assignment (Kasy and Sautmann, 2021). DOSE provides a flexible elicitation method that allows researchers to study such questions across a broad range of research environments and subject pools.

References

Abernethy, Jacob, Theodoros Evgeniou, Olivier Toubia, and Jean-Philippe Vert, "Eliciting Consumer Preferences Using Robust Adaptive Choice Questionnaires," *IEEE Transactions on Knowledge and Data Engineering*, 2008, *20* (2), 145–155.

Agranov, Marina, Andrew Schotter, and Isabel Trevino, "Complex for whom? an experimental approach to subjective complexity," *Available at agranov. caltech. edu*, 2025.

Aguiar, Mark, Mark Bils, and Corina Boar, "Who are the Hand-to-Mouth?," *Review of Economic Studies*, forthcoming.

Aigner, Dennis J., "A Brief Introduction to the Methodology of Optimal Experimental Design," *Journal of Econometrics*, 1979, *11* (1), 7–26.

Allcott, Hunt, Matthew Gentzkow, and Lena Song, "Digital Addiction," *American Economic Review*, 2022, *112* (7), 2424–2463.

Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutström, "Elicitation Using Multiple Price List Formats," *Experimental Economics*, 2006, *9* (4), 383–405.

_ , _ , _ , and _ , "Eliciting Risk and Time Preferences," *Econometrica*, 2008, *76* (3), 583–618.

Andersson, Ola, Håkan J. Holm, Jean-Robert Tyran, and Erik Wengström, "Risk Aversion Relates to Cognitive Ability: Preferences or Noise?," *Journal of the European Economic Association*, 2016, *14* (5), 1129–1154.

Andreoni, James and Charles Sprenger, "Estimating Time Preferences from Convex Budgets," *The American Economic Review*, 2012, *102* (7), 3333–3356.

_ , Michael A. Kuhn, and Charles Sprenger, "Measuring Time Preferences: A Comparison of Experimental Methods," *Journal of Economic Behavior & Organization*, 2015, *116*, 451–464.

_ , Michael Callen, Karrar Hussain, Muhammad Yasir Khan, and Charles Sprenger, "Using Preference Estimates to Customize Incentives: An Application to Polio Vaccination Drives in Pakistan," *Journal of the European Economic Association*, 2023, *21* (4), 1428–1477.

Apesteguia, Jose and Miguel A. Ballester, "Monotone Stochastic Choice Models: The Case of Risk and Time Preferences," *Journal of Political Economy*, 2018, *126* (1), 74–106.

Balietti, Stefano, Brennan Klein, and Christoph Riedl, "Optimal Design of Experiments to Identify Latent Behavioral Types," *Experimental Economics*, 2021, *24* (3), 772–799.

Barrington-Leigh, Christopher P., "The Econometrics of Happiness: Are we Underestimating the Returns to Education and Income?," *Journal of Public Economics*, 2024, *230*, 105052.

Benjamin, Daniel J., Sebastian A. Brown, and Jesse M. Shapiro, "Who Is 'Behavioral'? Cognitive Ability and Anomalous Preferences," *Journal of the European Economic Association*, 2013, *11* (6), 1231–1255.

Berry, James, Greg Fischer, and Raymond Guiteras, "Eliciting and Utilizing Willingness to Pay: Evidence from Field Trials in Northern Ghana," *Journal of Political Economy*, 2020, *128* (4), 1436–1473.

Bilbiie, Florin O., "Monetary Policy and Heterogeneity: An Analytical Framework," *Re-*

view of Economic Studies, forthcoming.

Booij, Adam S., Bernard M.S. Van Praag, and Gijs Van De Kuilen, “A Parametric Analysis of Prospect Theory’s Functionals for the General Population,” *Theory and Decision*, 2010, *68* (1-2), 115–148.

Burdea, Valeria and Jonathan Woon, “Online belief elicitation methods,” *Journal of Economic Psychology*, 2022, *90*, 102496.

Burfurd, Ingrid and Tom Wilkening, “Cognitive Heterogeneity and Complex Belief Elicitation,” *Experimental Economics*, 2022, *25* (2), 557–592.

Callen, Michael, Mohammad Isaqzadeh, James D. Long, and Charles Sprenger, “Violence and Risk Preference: Experimental Evidence from Afghanistan,” *The American Economic Review*, 2014, *104* (1), 123–48.

Camerer, Colin, Devdeepta Bose, Remi Daviet, and Taisuke Imai, “Social Preference Estimation Using Adaptive Experimental Design,” September 2020, (4899381).

Cavagnaro, Daniel R., Gabriel J. Aranovich, Samuel M. McClure, Mark A. Pitt, and Jay I. Myung, “On the Functional Form of Temporal Discounting: An Optimized Adaptive Test,” *Journal of Risk and Uncertainty*, 2016, *52* (3), 233–254.

_ , Jay I. Myung, Mark A. Pitt, and Janne V. Kujala, “Adaptive Design Optimization: A Mutual Information-Based Approach to Model Discrimination in Cognitive Science,” *Neural Computation*, 2010, *22* (4), 887–905.

_ , Mark A. Pitt, Richard Gonzalez, and Jay I. Myung, “Discriminating Among Probability Weighting Functions using Adaptive Design Optimization,” *Journal of Risk and Uncertainty*, 2013, *47* (3), 255–289.

_ , Richard Gonzalez, Jay I. Myung, and Mark A. Pitt, “Optimal Decision Stimuli for Risky Choice Experiments: An Adaptive Approach,” *Management Science*, 2013, *59* (2), 358–375.

Chakraborty, Anujit, Evan M. Calford, Guidon Fenig, and Yoram Halevy, “External and Internal Consistency of Choices made in Convex Time Budgets,” *Experimental Economics*, 2017, *20*, 687–706.

Chaloner, Kathryn and Isabella Verdinelli, “Bayesian Experimental Design: A Review,” *Statistical Science*, 1995, *10* (3), 273–304.

Chapman, Jonathan and Geoffrey Fisher, “Preference Elicitation: Common Methods and Potential Pitfalls,” in Erik Snowberg and Leeat Yariv, eds., *Handbook of Experimental Methodology*, Vol. 1 2025, chapter 2. Forthcoming.

_ , Erik Snowberg, Stephanie Wang, and Colin F. Camerer, “Loss Attitudes in the U.S. Population: Evidence from Dynamically Optimized Sequential Experimentation (DOSE),” 2018. NBER Working Paper #25072.

_ , _ , _ , and _ , “Looming Large or Seeming Small? Attitudes Towards Losses in a Representative Sample,” *Review of Economic Studies*, forthcoming.

_ , Mark Dean, Pietro Ortoleva, Erik Snowberg, and Colin Camerer, “Econographics,” *Journal of Political Economy Microeconomics*, 2023, *1* (1), 115–161.

_ , _ , _ , _ , and _ , “Willingness to Accept, Willingness to Pay, and Loss Aversion,” 2023. NBER Working Paper #30836.

_ , Pietro Ortoleva, Erik Snowberg, and Colin Camerer, “Time Stability of Behavioral (and other) Measures,” 2024. Mimeo.

_ , _ , _ , Leeat Yariv, and Colin Camerer, “Reassessing Qualitative Self-Assessments and Experimental Validation," 2024. Unpublished Working Paper.

Charness, Gary and Angelino Viceisza, "Three Risk-Elicitation Methods in the Field: Evidence from Rural Senegal," *Review of Behavioral Economics*, 2016, *3* (2), 145–171.

Clay, Summer N. Allison M. Harris, and Catherine L. Reed, "Loss Aversion Affects Inhibitory Processes for Reward and Loss as Indicated by Inhibition of Return," *Journal of Vision*, 2016, *16* (12), 95.

_ , John A. Clithero, Alison M. Harris, and Catherine L. Reed, "Loss Aversion Reflects Information Accumulation, Not Bias: A Drift-Diffusion Model Study," *Frontiers in Psychology*, 2017, *8*, 1708.

Condon, David M. and William Revelle, "The International Cognitive Ability Resource: Development and Initial Validation of a Public-Domain Measure," *Intelligence*, 2014, *43*, 52–64.

Crosetto, Paolo and Antonio Filippin, "A Theoretical and Experimental Appraisal of Four Risk Elicitation Methods," *Experimental Economics*, 2016, *19*, 613–641.

Dave, Chetan, Catherine C. Eckel, Cathleen A. Johnson, and Christian Rojas, "Eliciting Risk Preferences: When is Simple Better?," *Journal of Risk and Uncertainty*, 2010, *41* (3), 219–243.

Dohmen, Thomas, Armin Falk, Armin Huffman, and Uwe Sunde, "Are Risk Aversion and Impatience Related to Cognitive Ability?," *The American Economic Review*, 2010, *100* (3), 1238–1260.

_ , _ , David Huffman, and Uwe Sunde, "On the Relationship between Cognitive Ability and Risk Preference," *Journal of Economic Perspectives*, 2018, *32* (2), 115–34.

_ , _ , _ , _ , Jürgen Schupp, and Gert G. Wagner, "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences," *Journal of the European Economic Association*, 2011, *9* (3), 522–550.

Drake, Marshall, Fernando Payró, Neil Thakral, and Linh T Tô, "Bayesian Adaptive Choice Experiments," Technical Report, Mimeo 2022.

Echenique, Federico and Siddharth Prasad, "Incentive Compatible Active Learning," *arXiv preprint arXiv:1911.05171*, 2019.

Eckel, Catherine C. and Philip J. Grossman, "Sex Differences and Statistical Stereotyping in Attitudes toward Financial Risk," *Evolution and Human Behavior*, 2002, *23* (4), 281–295.

_ and _ , "Men, Women and Risk Aversion: Experimental Evidence," in Charles R. Plott and Vernon L. Smith, eds., *Handbook of Experimental Economics Results*, Vol. 1, North-Holland, 2008, pp. 1061–1073.

El-Gamal, Mahmoud A. and Thomas R. Palfrey, "Economical Experiments: Bayesian Efficient Experimental Design," *International Journal of Game Theory*, 1996, *25* (4), 495–517.

El-Gamal, Mahmoud A., Richard D. McKelvey, and Thomas R. Palfrey, "A Bayesian Sequential Experimental Study of Learning in Games," *Journal of the American Statistical Association*, 1993, *88* (422), 428–435.

Fagereng, Andreas, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri, "Heterogeneity and Persistence in Returns to Wealth," *Econometrica*, 2020, *88* (1), 115–170.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde, "Global Evidence on Economic Preferences," *The Quarterly Journal* *of Economics*, 2018, pp. 1645—-1692.

_ , _ , _ , **David Huffman, and Uwe Sunde**, "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences," *Management Science*, 2023, *69* (4), 1935–1950.

Filippin, Antonio and Paolo Crosetto, "A Reconsideration of Gender Differences in Risk Attitudes," *Management Science*, 2016, *62* (11), 3138–3160.

Frederick, Shane, "Cognitive Reflection and Decision Making," *Journal of Economic Perspectives*, 2005, *19* (4), 25–42.

Freundt, Jana, Holger Herz, and Leander Kopp, "Intrinsic Preferences for Autonomy," 2023. CESifo Working Paper # 10,342.

Frey, Renato, Andreas Pedroni, Rui Mata, Jörg Rieskamp, and Ralph Hertwig, "Risk Preference Shares the Psychometric Structure of Major Psychological Traits," *Science Advances*, 2017, *3* (10), e1701381.

Frydman, Cary, Colin Camerer, Peter Bossaerts, and Antonio Rangel, "MAOA-L Carriers are Better at Making Optimal Financial Decisions under Risk," *Proceedings of the Royal Society of London B: Biological Sciences*, 2011, *278* (1714), 2053–2059.

Fuster, Andreas, Greg Kaplan, and Basit Zafar, "What Would you do with $500? Spending Responses to Gains, Losses, News, and Loans," *The Review of Economic Studies*, 2021, *88* (4), 1760–1795.

Gideon, Michael, Joanne Hsu, and Brooke Helppie-McFall, "Heaping at Round Numbers on Financial Questions: The Role of Satisficing," *Survey Research Methods*, 2017, *11* (22), 189–214.

Gillen, Ben, Erik Snowberg, and Leeat Yariv, "Experimenting with Measurement Error: Techniques and Applications from the Caltech Cohort Study," *Journal of Political Economy*, 2019, *127* (4), 1826–1863.

Gneezy, Uri and Jan Potters, "An Experiment on Risk Taking and Evaluation Periods," *The Quarterly Journal of Economics*, 1997, *112* (2), 631–645.

Golovin, Daniel and Andreas Krause, "Adaptive Submodularity: Theory and Applications in Active Learning and Stochastic Optimization," *Journal of Artificial Intelligence Research*, November 2011, *42*, 427–486.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, "Time varying risk aversion," *Journal of Financial Economics*, 2018, *128* (3), 403–421.

Hey, John D., Andrea Morone, and Ulrich Schmidt, "Noise and Bias in Eliciting Preferences," *Journal of Risk and Uncertainty*, 2009, *39*, 213–235.

Hryshko, Dmytro, María José Luengo-Prado, and Bent E Sørensen, "Childhood determinants of risk aversion: The long shadow of compulsory education," *Quantitative Economics*, 2011, *2* (1), 37–72.

Imai, Taisuke and Colin F. Camerer, "Estimating Time Preferences from Budget Set Choices Using Optimal Adaptive Design," 2018. Ludwig Maximilian University of Munich, *mimeo.*

Jacobson, Sarah and Ragan Petrie, "Learning from Mistakes: What Do Inconsistent Choices Over Risk Tell Us?," *Journal of Risk and Uncertainty*, 2009, *38* (2), 143–158.

Jagelka, Tomáš, "Are Economists' Preferences Psychologists' Personality Traits? A Structural Approach," *Journal of Political Economy*, 2024, *132* (3), 910–970.

Johnson, Cathleen, Aurélien Baillon, Han Bleichrodt, Zhihua Li, Dennie Van

Dolder, and Peter P Wakker, "PRINCE: An Improved Method for Measuring Incentivized Preferences," *Journal of Risk and Uncertainty*, 2021, *62* (1), 1–28.

Kahneman, Daniel and Amos Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 1979, *47* (2), 263–291.

Kasy, Maximilian and Anja Sautmann, "Adaptive Treatment Assignment in Experiments for Policy Choice," *Econometrica*, 2021, *89* (1), 113–132.

Kirby, Kris N., "One-year Temporal Stability of Delay-Discount Rates," *Psychonomic Bulletin & Review*, 2009, *16* (3), 457–462.

Krajbich, Ian, Colin Camerer, and Antonio Rangel, "Exploring the Scope of Neurometrically Informed Mechanism Design," *Games and Economic Behavior*, 2017, *101*, 49–62.

Krosnick, Jon A. and Stanley Presser, "Question and Questionnaire Design," in Peter V. Marsden and James D. Wright, eds., *Handbook of Survey Research*, 2010, pp. 263–313.

Kujala, Janne V. and Tuomas J. Lukka, "Bayesian Adaptive Estimation: The Next Dimension," *Journal of Mathematical Psychology*, 2006, *50* (4), 369–389.

Lesmes, Luis Andres, Seong-Taek Jeon, Zhong-Lin Lu, and Barbara Anne Dosher, "Bayesian Adaptive Estimation of Threshold Versus Contrast External Noise Functions: The Quick TvC Method," *Vision Research*, 2006, *46* (19), 3160–3176.

Levin, Irwin P., Stephanie S. Hart, Joshua A. Weller, and Lyndsay A. Harshman, "Stability of Choices in a Risky Decision-Making task: A 3-year Longitudinal Study with Children and Adults," *Journal of Behavioral Decision Making*, 2007, *20* (3), 241–252.

Lewi, Jeremy, Robert Butera, and Liam Paninski, "Sequential Optimal Design of Neurophysiology Experiments," *Neural Computation*, 2009, *21* (3), 619–687.

Lönnqvist, Jan-Erik, Markku Verkasalo, Gari Walkowitz, and Philipp C. Wichardt, "Measuring Individual Risk Attitudes in the Lab: Task or Ask? An Empirical Comparison," *Journal of Economic Behavior & Organization*, 2015, *119*, 254–266.

Mata, Rui, Renato Frey, David Richter, Jürgen Schupp, and Ralph Hertwig, "Risk Preference: A View from Psychology," *Journal of Economic Perspectives*, 2018, *32* (2), 155–72.

McFadden, Daniel, "Economic Choices," *The American Economic Review*, 2001, *91* (3), 351–378.

Meier, Stephan and Charles Sprenger, "Temporal Stability of Time Preferences," *Review of Economics and Statistics*, 2015, *97* (2), 273–286.

Mitani, Yohei and Nobuyuki Hanaki, "Pay a lot to a few instead of a bit to all! Evidence from online donation experiments," 2025. Institute of Social and Economic Research, Osaka University Discussion Paper 1273.

Moffatt, Peter, "Optimal Experimental Designs in Models of Decision and Choice," in Marcel Boumans, ed., *Measurement in Economics: A Handbook*, Elsevier, 2007.

Müller, Peter, Don A. Berry, Andy P. Grieve, Michael Smith, and Michael Krams, "Simulation-Based Sequential Bayesian Design," *Journal of Statistical Planning and Inference*, 2007, *137* (10), 3140–3150.

Paninski, Liam, "Asymptotic Theory of Information-Theoretic Experimental Design," *Neural Computation*, 2005, *17* (7), 1480–1507.

Pedroni, Andreas, Renato Frey, Adrian Bruhin, Gilles Dutilh, Ralph Hertwig,

and Jörg Rieskamp, "The Risk Elicitation Puzzle," *Nature Human Behaviour*, 2017, *1* (11), 803.

Peirce, Charles S., "Note on the Theory of the Economy of Research," *United States Coast Survey for the Fiscal Year Ending June 1876, US Government Printing Office 1879, reprinted in Operations Research*, 1967, *15*, 643–648.

Pew Research Center, *Evaluating Online Nonprobability Surveys*, www.pewresearch.org, May 2016.

Preuss, Malte, "Intra-individual Stability of Two Survey Measures on Forward-looking Attitude," *Journal of Economic Behavior & Organization*, 2021, *190*, 201–227.

Ray, Debajyoti, Daniel Golovin, Andreas Krause, and Colin Camerer, "Bayesian Rapid Optimal Adaptive Design (BROAD): Method and Application Distinguishing Models of Risky Choice," 2012. Caltech, *mimeo.*

Rubinstein, Ariel, "A Typology of Players: Between Instinctive and Contemplative," *The Quarterly Journal of Economics*, 2016, *131* (2), 859–890.

Ryan, Elizabeth G. Christopher C. Drovandi, James M. McGree, and Anthony N. Pettitt, "A Review of Modern Computational Algorithms for Bayesian Optimal Design," *International Statistical Review*, 2016, *84* (1), 128–154.

Schildberg-Hörisch, Hannah, "Are Risk Preferences Stable?," *Journal of Economic Perspectives*, 2018, *32* (2), 135–54.

Schwarz, Norbert and Daphna Oyserman, "Asking Questions about Behavior: Cognition, Communication, and Questionnaire Construction," *The American Journal of Evaluation*, 2001, *22* (2), 127–160.

Sokol-Hessner, Peter, Ming Hsu, Nina G. Curley, Mauricio R. Delgado, Colin F. Camerer, and Elizabeth A. Phelps, "Thinking Like a Trader Selectively Reduces Individuals' Loss Aversion," *Proceedings of the National Academy of Sciences*, 2009, *106* (13), 5035–5040.

Stango, Victor and Jonathan Zinman, "Behavioral Biases are Temporally Stable," 2020.

_ and _ , "We Are All Behavioral, More or Less: A Taxonomy of Consumer Decision Making," *Review of Economic Studies*, 2023, *90* (3), 1470–1498.

Taubinsky, Dmitry and Alex Rees-Jones, "Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment," *The Review of Economic Studies*, 2018, *85* (4), 2462–2496.

Toubia, Olivier, Eric Johnson, Theodoros Evgeniou, and Philippe Delquié, "Dynamic Experiments for Estimating Preferences: An Adaptive Method of Eliciting Time and Risk Parameters," *Management Science*, 2013, *59* (3), 613–640.

_ , John R. Hauser, and Duncan I. Simester, "Polyhedral Methods for Adaptive Choice-Based Conjoint Analysis," *Journal of Marketing Research*, 2004, *41* (1), 116–131.

von Gaudecker, Hans-Martin, Arthur Van Soest, and Erik Wengström, "Heterogeneity in Risky Choice Behavior in a Broad Population," *The American Economic Review*, 2011, *101* (2), 664–694.

Wald, Abraham, "Foundations of a general theory of sequential decision functions," *Econometrica, Journal of the Econometric Society*, 1947, pp. 279–313.

_ , *Statistical Decision Functions*, John Wiley and Sons, New York; Chapman and Hall, London., 1950.

Wang, Stephanie W. Michelle Filiba, and Colin Camerer, "Dynamically Optimized

Sequential Experimentation (DOSE) for Estimating Economic Preference Parameters,” 2010. Caltech, *mimeo.*

Online Appendix—Not Intended for Publication

Table of Contents

A Risk and Time Preferences in Prior Studies

In this appendix we compare our results regarding the correlates and temporal stability of economic preferences to those from prior studies. We start by comparing the estimates of risk and time preference in different samples to the prior literature, and then—building on the main text—discuss prior studies investigating the correlates and temporal stability of those preferences.

A.1 DOSE Estimates in Different Samples

Our findings regarding the extent of risk aversion and discounting in the general population are broadly similar to those of previous studies in representative populations; the few differences appear to be explained by our elicitation method. Risk aversion and discounting are widespread amongst our survey participants, but the median level of risk aversion is lower than found in the previous literature—a difference that may be explained by our use of binary choice questions rather than MPLs.

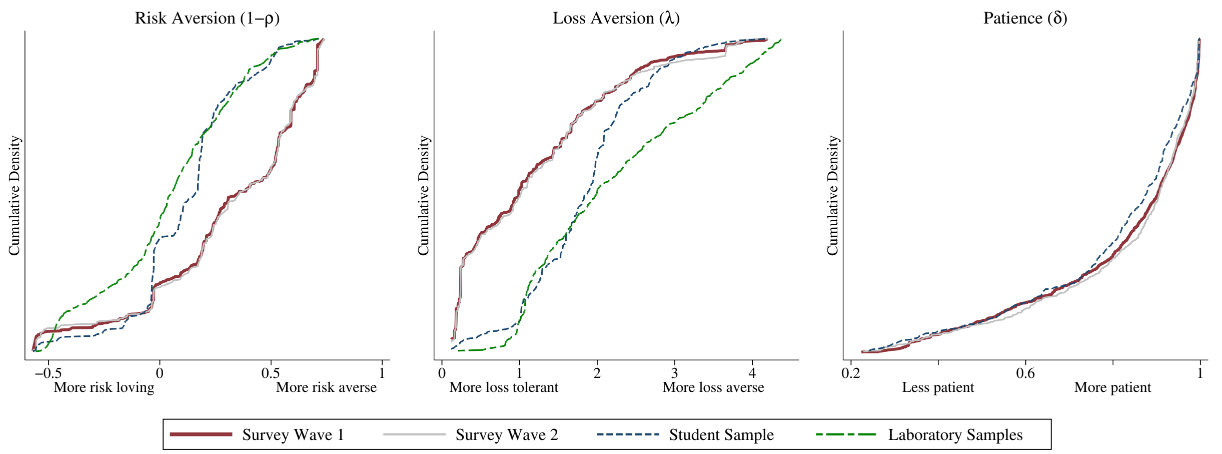
In addition to our general population survey, we also use DOSE to elicit risk and time preferences from undergraduate students both online and in the laboratory, providing a clearer comparison to the many studies investigating preferences in undergraduate populations.

Online Student Sample: We implemented a 10-question risk-loss DOSE module in an online survey of undergraduate students ($N = 369$) recruited from the University of Pittsburgh Experimental Laboratory (PEEL). The survey was very similar to our general population survey. As with our representative sample: participants completed the survey through YouGov's online portal and questions were presented with the same point value. The main difference between the samples is that the students received payment in cash, via Visa gift card, rather than in YouGov points.[1]

Laboratory Samples: We compare our student sample to a sample ($N = 439$) of participants completing DOSE modules in four previous laboratory studies. Each of these studies implemented DOSE (based on our original working paper, Wang et al. (2010)) to estimate risk and loss aversion using binary choices with the same structure as those in our survey.

[1] Payment was completed within two weeks of survey completion using an exchange rate of 1,000 points =\$1. The survey was completed in January 2019. The median time to complete the survey was shorter than among our representative sample—22 minutes—partly because there were fewer tasks, and partly because the students went through each task an average of 25% faster. The average payment was \$15.50 due to the removal of some low-average-value tasks.

Figure A.1: The general population is more risk tolerant and loss averse than students.



Notes: The figure displays the kernel density of each parameter, plotted using Epanechnikov kernel with bandwidth chosen by rule-of-thumb estimator. "Student Sample" refers to students completing DOSE in an online survey. "Laboratory Samples" refers to students completing DOSE in previous laboratory experiments—see text above for details.

We take those individual choices, and then re-estimate the individual risk and loss aversion parameters using the same procedure and priors as used for our main estimates.[2]

Differences in elicitation method appear to explain the lower level of risk aversion estimated by DOSE than found in previous studies of representative samples. The mean and median Coefficient of Relative Risk Aversion (1-ρ) are 0.30 and 0.34 respectively, compared to previous findings ranging from approximately 0.4 (Dohmen et al., 2010) to 0.7 (Harrison et al., 2007; Andersen et al., 2008). This pattern is consistent with laboratory studies finding lower levels of risk aversion using binary choice questions: the median coefficient using DOSE is 0.05 in the lab and 0.17 in our student survey—similar to the value of 0.12 found by Sokol-Hessner et al. (2009) using binary questions, but much lower than the range of 0.3–0.5 found by Holt and Laury (2002). Moreover, the median coefficients on the MPLs on our survey (0.4 and 2.1) are more in line with previous studies.

For discounting, we estimate a median monthly discount factor (δ) of 0.90, which is similar to the monthly discount factor reported by Meier and Sprenger (2010), and falls in the lowest quartile of the results of three recent laboratory studies using the Convex Time Budget method of Andreoni and Sprenger (2012)—see Imai and Camerer (2018, Appendix Table D1). In our student sample, we find a slightly lower median discount factor of 0.87—however given the differences in payment across samples (cash versus points), it is difficult

[2]58 participants completed DOSE at Caltech (Krajbich et al., 2017), 172 at Claremont McKenna (108 collected by Clay et al. 2017 and 64 by Clay et al. 2016), and 209 students at UCLA (this data was generously provided to us by Alec Smith).

to compare directly.

The results in Figure A.1 provide further evidence that DOSE is an effective tool in online surveys. We find that students are both less risk averse and more loss averse than the general population, whether DOSE is used in the laboratory or in an online survey—the survey implementation does not appear to affect our results. Second, the distributions are very similar across survey waves, suggesting that participants make consistent choices over time—it does not appear that initial confusion or attempts to "game" DOSE during the second elicitation are a major factor in the estimates.

A.2 Correlates of Economic Preferences

The patterns of correlation between economic preferences and sociodemographic characteristics we find (see Table 2), largely match the literature. The relationship between cognitive ability and patience, unlike risk aversion, is well-established in both economics and psychology (for example, Shamosh and Gray, 2008; Burks et al., 2009; Dohmen et al., 2010; Benjamin et al., 2013). Patient individuals have also been found to have higher income, greater savings, and more education (see DellaVigna and Paserman, 2005; Urminsky and Zauberman, 2016; Falk et al., 2018). In the laboratory most studies have documented that women are more risk averse than men (Eckel and Grossman, 2008), as have Falk et al. (2018) in a representative sample. There is also some evidence of a negative relationship between risk aversion and income, although results have been mixed (see, for instance Dohmen et al., 2010; Barsky et al., 1997). A recent meta-analysis finds, in line with our results, that age is not related to risk aversion, but is related to time preference (Bagaïni et al., 2023). Also consistent with our results, Olschewski et al. (2018) find that greater cognitive load leads to less consistent choice.

The most important difference between our results and the literature is the relationship between risk aversion and cognitive ability, where we find a strong negative correlation, whereas results in the previous literature are mixed between somewhat smaller negative correlations, no correlations and, occasionally positive correlations (see Lilleholt, 2019, for a metaanalysis). As discussed in Section 5.3, it appears this difference is explained by the fact DOSE accounts for inconsistent choice.

A.3 Temporal Stability of Economic Preferences

As noted in Section 5.4, the over-time correlation of the DOSE preference estimates is larger than estimates in most previous studies, but direct comparisons are complicated by differences in the sample used. Chuang and Schechter (2015) review previous studies of stability of risk and/or time preferences, reporting that most studies use hypothetical methods. A few incentivized studies report higher correlations than DOSE over much shorter periods of time (Wölbert and Riedl, 2013; Dean and Sautmann, 2014; Falk et al., 2023). Mata et al. (2018) provide a more recent meta-analysis, and find that over a five-year period the stability of lottery-based risk elicitations—both incentivized and hypothetical—is 0.2. In another field study, Kirby et al. (2002) reports correlations of 0.09–0.23 over a six month period among Bolivian Amerindians.

The variety of the methodologies, both in terms of time between elicitations, and in sample composition — for instance, it is not clear how best to compare our representative sample to selected sub-samples such as Meier and Sprenger (2015)'s low-income Americans—mean that we cannot identify the reasons for differences between studies. However, in general, we identify higher stability than most previous estimates, particularly outside the laboratory—consistent with the DOSE estimates being less affected by attenuation bias due to measurement error.

B DOSE Procedure and Survey Implementation

This Appendix discusses the potential for strategic manipulation of DOSE, and presents details of our survey implementation.

B.1 The Potential for Gaming DOSE and Incentive Compatibility

In this Appendix we discuss the potential issue of gaming in DOSE and other dynamic elicitation methods. While in principle a participant could have an incentive to game adaptive question sequences, we argue, drawing on previous work, that this is unlikely to be a concern in practice.

Let's start with a scenario which is the easiest case for gaming and hence is the worst case from the point of view of an experimenter. A participant faces a series of simple choices between a certain payment \$X or a 50% coin flip between 0 and \$10. The value of X is varied dynamically to pin down a certainty-equivalent amount which is equal to the gamble value. Suppose the initial choice is between the gamble and \$X=4. A person who truly prefers \$4 might strategically choose the gamble, if they think that doing so would lead the experimenter to increase the values of \$X in subsequent questions to "titrate" how much they tolerate risk. Expressing an exaggerated preference for risk, by turning down \$4, could benefit them over their entire portfolio of immediate and—hypothesized—later choices that are dynamically-adjusted. There is no question that the possibility of dynamic gaming exists and could in principle undermine inference. For example, if the hypothetical subject above keeps turning down gambles in order to get better and better ($X; coin flip) binary choices she will earn more money, and her Bayesian posterior will put too much weight on risk-preference.

However, in practice, gaming of this kind will be a major concern only if a number of quite stringent conditions hold. In particular, gaming requires all of following: (1) participants believe that future test choices depend on previous responses (recall that they are not informed of this fact in our survey); (2) participants can and do compute how to misrepresent preferences in early choices to create better future choices (as evaluated by their own preferences); and (3) the expected value of misrepresentation computed in step (2) must be high enough to be worthwhile, ex ante. In a "rational gaming" type analysis, if they don't believe in (1) then they won't proceed to (2). If step (3) isn't satisfied they will not make the step (2) computation. These conditions require a high level of participant sophistication, ability to undertake complex calculations, and understanding of the DOSE algorithm. Moreover, they require that gaming is actually beneficial, which is not necessarily the case, as discussed further below.

Ray et al. (2012) provide direct evidence regarding steps (2) and (3) of this chain. First, to test whether participants actually game adaptive algorithms (condition (2)) and what happens if they try to, they compare a baseline control with an augmented-instructions treatment in which they explicitly instructed subjects that gaming could help them. Specifically, the instructions said (excerpted):

> It could help you to know that an adaptive algorithm generates the choices between two lotteries that are presented to you... It is possible, mathematically, to choose lotteries you do not like in the beginning in order to lead the adaptive algorithm to create future lottery choices that you like better... If you think you can guide the algorithm in this way, by choosing lotteries you don't like to create better choice pairs in future trials, you are free to do so.

The results showed little evidence that gaming was important. The distributions of best-fit risky choice models did not differ in the control and explicitly-instructed treatment. Further, in post-experiment debriefings, a majority of participants said they did not consider gaming. This claim is supported by direct evidence of (a lack of) gaming in their data. Specifically, Ray et al. (2012) point out that gaming should leave two clear fingerprints. First, response times will be longer when gaming is occurring, and drop sharply when gaming is finished. However, the response time profiles across trials in the experimental session were statistically indistinguishable between control and explicitly-instructed subjects, and followed a very regular learning-curve reduction in RT. Second, trial-by-trial parametric classifications should reverse from early to late trials if there is gaming. Recall our simple example of a risk-averse subject who is truly indifferent between \$X=\$4 and a \$0-10 coin flip, but pretends to be risk-seeking in early choices in order to get future choices with higher values of \$X. Their estimated risk-aversion parameter will go sharply down over early trials, then reverse and go sharply up. They rarely observed this type of parametric reversal in their data (see also Appendix Figure E.2 and surrounding discussion).

To contribute evidence about step 3, Ray et al. (2012) compute the expected gain from gaming, and find it is relatively low. Their example is a hypothetical person in a binary risky choice paradigm with (assumed) expected-value preferences who could perfectly compute optimal gaming for a 10-trial sequence of choices. This hypothetical person could access the "God screen" of how the DOSE sequence was created to pick binary choices in round t+1 based on the history of choices before t+1, and could perfectly compute how to game early choices to maximize expected payoff across all 10 trials (assuming all trials are equally likely to be chosen for payment). Under this best-case scenario, the hypothetical optimally-gaming agent earned 8% more. As such, participants had limited financial incentive to engage in gaming.[3]

In sum, gaming is not in practice rewarding, does not seem to be widespread, and is not evident either in response time patterns or classification reversals. As such, it is unlikely to be a concern in most settings, or pose a threat to our results.

Further, if gaming is a concern in other research contexts, then it is straightforward to implement DOSE with an incentive-compatible payment scheme. One method (Johnson et al., 2021) is to choose large design space, and sample one binary choice from that space. If the sampled choice was in the set of sequential choices that were actually faced, choose what the subject already picked. If it's a new choice, then their novel choice will count for pay. This method removes the gaming incentive, by breaking the link between the sequence of questions asked and the likelihood a question is chosen for payment. Another method, suggested by Ian Krajbich, and implemented in Krajbich et al. (2017), is to use the sampled choices as a way of training an algorithm—the parametric specification which is encoded in

[3]Informal theory offers an explanation for the low benefit to gaming. Specifically, in later periods, it does not pay to strategize since doing so makes suboptimal immediate choices. In early periods, strategizing is immediately costly (for the same reason), but may offer benefits in future questions—there is a natural tradeoff between the cost of strategizing in early periods, and possible future gains. In a sequence of ten questions the short-run costs are relatively large versus these future gains. While gaming could be more of a concern in a longer DOSE sequence, this discussion also suggests an easy remedy: exclude some fraction of earlier periods and recompute Bayesian posteriors using only later periods, or adjust the randomization weights of choosing a single choice to resolve for pay so later choices are much more heavily weighted.

the posterior from the last actual choice. The trained choice specification is then used to make a novel choice. In this method, the subjects are essentially "training" an algorithm, much as choices of Amazon books are training a recommender system.

B.2 DOSE in a Representative Survey

We now turn to the practical details of implementing DOSE in two waves of a large, representative, incentivized survey of the U.S. population. The survey includes two DOSE modules—one relating to risk and one to time preferences—as well as other behavioral elicitations, and cognitive and sociodemographic questions.

B.2.1 DOSE Modules

Both of the DOSE modules were comprised of ten questions selected using the procedure described in Section 2, slightly modified due to the practicalities associated with implementing the survey online. The design of YouGov's online platform precluded using DOSE to choose questions in real time and so, instead, simulated responses were used to map out all possible sets of binary choices in advance. That tree was then used to route participants through the survey. Mapping such a tree with a refined prior was infeasible given both computational constraints and the limitations of YouGov's interface (mapping such a tree over 20 questions would involve over 500,000 routes through the survey). As such, questions were selected using a coarser prior and then final individual-level estimates were obtained by performing the Bayesian updating procedure with a joint 100-point discretized uniform prior.[4]

We now describe the particular design choices—the priors, utility specification, and question set—that we made for each of the two DOSE modules.

Risk Preferences: The first DOSE module elicited risk and loss aversion. Participants were given 10,000 points and offered a sequence of ten binary choices between a 50:50 lottery and a sure amount. Two types of lottery were used. The first had a 50% chance of 0 points, and a 50% chance of winning a (varying) positive amount of points (of up to 10,000). The second had a 50% chance of winning an amount up to 10,000 points, and a 50% chance of a loss of up to 10,000 points. In the latter case, the sure amount was always 0 points.[5]

[4] The prior used for question selection included 12 mass points for ρ and δ, 20 for λ, and 4 for μ. To utilize the information about the curvature of the utility function from the risk-loss module, for the time preference module participants were assigned to one of ten prior distributions over ρ, based on their estimated ρ from the risk-loss module.

[5] The set of potential questions used for the risk module allowed for gains ranging between 1,000 and 10,000 points in 500 point increments, and sure amounts and losses varying ranging from 500 points to 10,000 points in 100 point increments. Questions were excluded if one choice was first-order stochastically

The question sequence for the risk preference module was selected using the specification in Equation (1) and a prior constructed using the estimates for laboratory participants obtained by Sokol-Hessner et al. (2009) and Frydman et al. (2011): 0.2–1.7 for ρ, and 0–4.6 for λ. To account for the survey environment we restricted the question selection procedure in two ways. First, to focus the procedure on obtaining a precise estimate of ρ before moving onto estimates of λ, the first four questions in the module were restricted to be lotteries over gains. Second, to make it harder for participants to identify the adaptive nature (and hence attempt to manipulate) the procedure, the maximum prize was restricted to be no more than 7,000 points in each even numbered round.

Time Preferences: The second DOSE module elicited discount factors and refined the estimates of the curvature of the utility function elicited in the risk preference module. Participants were offered a sequence of ten binary choices between a lower amount of points at an earlier date (either the day of the survey, or in the future) or a higher amount at a later date (up to 90 days in the future). The maximum payment in each question was 10,000 points.[6]

The time discounting questions were selected accounting for both discounting and present bias. As with the risk module, some restrictions were placed on the question selection procedure. The first five questions were restricted to the choice between payment on two dates in the future. In addition, when considering two options in the future (that is, $t_1 > 0$ and $t_2 > 0$), individuals were assumed to choose as if they have a fixed value of the present bias parameter (β=0.64, based on the estimates from Tanaka et al. (2010)).

B.2.2 Additional Measures

We also utilize survey measures of cognitive ability and more standard elicitations of risk and time preferences.

Cognitive Ability: Cognitive ability was measured using a set of nine questions: six from the International Cognitive Ability Resource (ICAR, Condon and Revelle, 2014) and three from the Cognitive Reflection Test (CRT) developed by Frederick (2005).

---

dominated for all values of the prior distribution. Questions were also selected as if the prize amounts were 3 times the actual amounts offered in the lottery to improve discrimination of the risk and loss aversion parameters.

[6]Possible payment amounts in the time module ranged from 1,000 to 10,000 points in 1,000 point increments, and possible payment days were of 0, 1, 3, 5, 7, 9, 10, 12, 16, 21, 28, 35, 42, 49, 56, 60, 70, 80, 90 days after the survey. The question set included all combinations of these payment amounts and dates in which the early payment was less than the later payment.

The IQ test consisted of three questions from the ICAR matrix module and three from the 3-D rotational module. In the matrix questions, participants were presented with a 3x3 matrix with 8 geometric designs. They then had to choose the correct design to complete the pattern from a list of 6 possible options. In the 3-D rotational questions, participants were shown a picture of three sides of a cube. They then had to choose which of six options (each also showing three sides of a cube) was compatible with a rotated version of the original.

The CRT includes three arithmetically straightforward questions with an instinctive, but incorrect, answer. The test thus measures the tendency for an individual to reflect upon a question rather than answer instinctively.

Risk Aversion MPLs: Two MPLs asked participants to choose between a fixed 50/50 lottery and a series of ascending sure amounts. The row in which the participant first chose the sure amount identified a range of possible certainty equivalents for the lottery—we use the midpoint of this range. There were two MPLs of this type: the first had a 50/50 lottery over 0 and 10,000 points, the second, a 50/50 lottery over 2,000 and 8,000 points.[7]

Time Preference MPLs: In addition to the DOSE module, the survey included two MPLs to elicit time preferences. The first time MPL elicited the amount of points that the participant valued the same as 6,000 points 45 days later. The second MPL elicited the amount of points in 45 days that the participant valued the same as 6,000 points in 90 days. We estimate monthly discount rates based on the midpoint of the row in which a participant switched to preferring money to today over the future. The bottom row of the MPLs allowed participants to specify they would prefer receiving 6,000 points in the future to receiving 6,100 points today. In this case they were assigned a discount rate of 1.

[7] These MPLs were designed to elicit the willingness to pay (WTP) for a lottery; participants were endowed with a fixed set of points and asked how much they would exchange for the lottery. An additional two MPLs in the survey elicited the willingness to accept (WTA) for the same lotteries; that is, participants were endowed with the lottery and asked the amount of points that they would need to be willing to exchange the lottery. See Chapman et al. (2023) for more details of the design of these measures, as well as extensive discussion of the relationship between WTA, WTP, and other risk elicitations in the survey.

C Simulations using Laboratory Choices

In this appendix we demonstrate the benefits of DOSE's personalized question sequence using data from 120 subjects in two prior laboratory experiments.[8] In each of these experiments, participants were asked the same set of 140 binary choices over gains and losses—with each choice having the same structure as the questions used in our survey. A 20-question DOSE procedure obtains parameter estimates that are close to (within 15% of) parameter estimates after 140 questions. Further, a joint uniform prior over the parameters is close to optimal for question selection. The Bayesian procedure also requires many fewer questions to elicit individual-specific estimates of risk and loss aversion than Maximum Likelihood Estimation.

C.1 Estimates of Risk and Loss Aversion

In our simulation, we optimally order the questions for each participant using DOSE and compare the parameter estimates to those that would be obtained under a random question ordering. After DOSE selects a question, we provide it with the answer the participant gave in the experiment. The procedure then updates the probability distribution over parameters, selects the next question, and so on. This allows us to compare, question by question, the *inaccuracy*—the absolute distance from the true parameter value as a percentage of the true value—of DOSE's estimates with those elicited by a random question ordering. As we do not have access to true parameter values, we substitute the values one would obtain using the choices in all 140 questions.

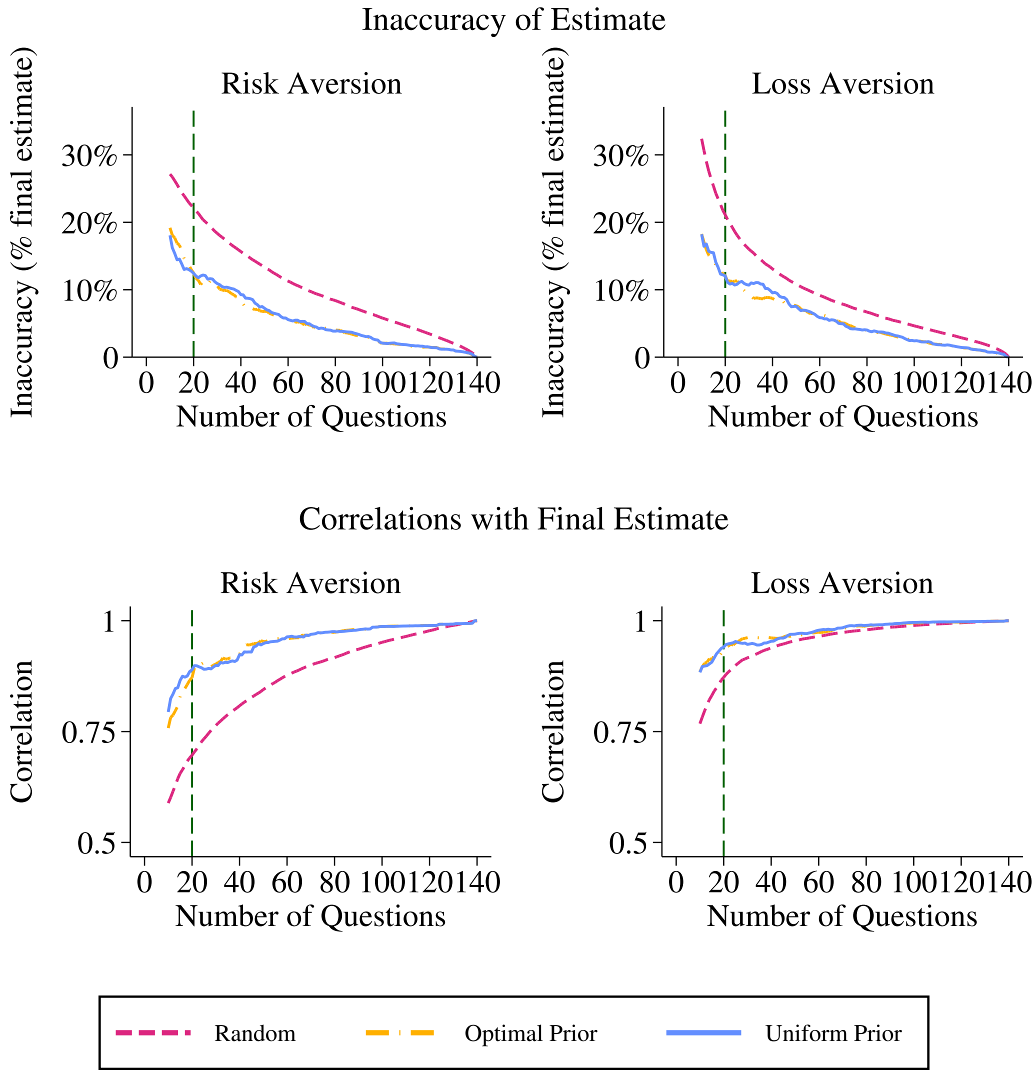
A 20-question DOSE sequence provides a similar amount of information as about 50 randomly-ordered questions, as shown in the top panel of Figure C.1.[9] The DOSE estimates of both risk and loss aversion are consistently closer to the final parameter estimate, indicating—under the assumption that the final estimate closely approximates an individual's true parameters—that the procedure provides accurate estimates considerably faster than selecting questions at random. After 20 questions, the DOSE estimates are almost twice as close to the final estimate as those under a random question ordering (12% vs. 21–22%).

The bottom panel of Figure C.1 shows that the DOSE estimates are also more highly correlated with the final estimates, an important feature when seeking to identify correlations between preferences and other population characteristics. After a 20-question DOSE sequence, the correlations are higher than obtained under the random ordering for both loss aversion (0.94 versus 0.87) and risk aversion (0.89 versus 0.70).

[8] 90 participants come from Frydman et al. (2011) and 30 from Sokol-Hessner et al. (2009).

[9] For loss aversion, 45 randomly-ordered questions are needed to be as close to the final estimate as 20 DOSE questions. For risk aversion, 55 questions are required.

Figure C.1: Optimal question selection rapidly leads to accurate estimates.

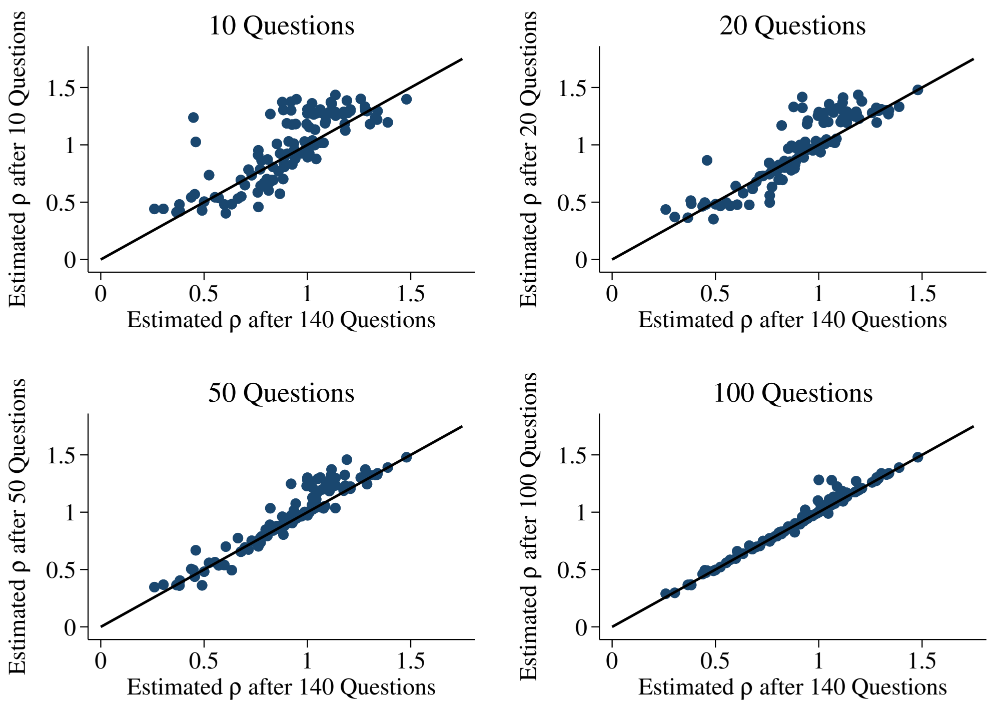


Notes: Based on data from Sokol-Hessner et al. (2009) and Frydman et al. (2011). “Optimal Prior” and “Uniform Prior” refer to DOSE question selection using corresponding priors. “Random” orders questions randomly, averaging over 100 different random orderings. In the top panel, each line shows the inaccuracy of Bayesian estimates obtained after each question (starting at question 10). The bottom panel displays the correlation between the Bayesian estimates obtained after each question and the final estimate. All parameter values are estimated using a uniform initial prior.

These simulations also show that using the uniform prior is close to optimal for question selection. We compare the performance of DOSE question selection using a uniform prior to that using an *optimal prior* constructed from the distribution of the estimates after 140 questions. To focus on the question selection impacts of the prior, we estimate the parameter values using a uniform prior in both cases. As shown in Figure 3, both the accuracy and the correlations are similar whether using the optimal or uniform prior.

DOSE improves the accuracy of individual-specific estimates for the full distribution of parameter values, as shown in Figures C.2 and C.3. These figures display the progression towards the final parameter estimates after asking 10, 20, 50 and 100 DOSE questions. After just 10 questions, the estimates for both risk and loss aversion are clustered around the 45 degree line, reflecting a high degree of correlation with the final estimates. There is no evidence for any of the parameters that the procedure converges faster for particular parameter values: indicating that accuracy improvements from DOSE are not an artefact of the particular distribution of preference parameters we observe.

Figure C.2: Correlations between final estimates of the risk aversion parameter and the estimates after selected rounds.

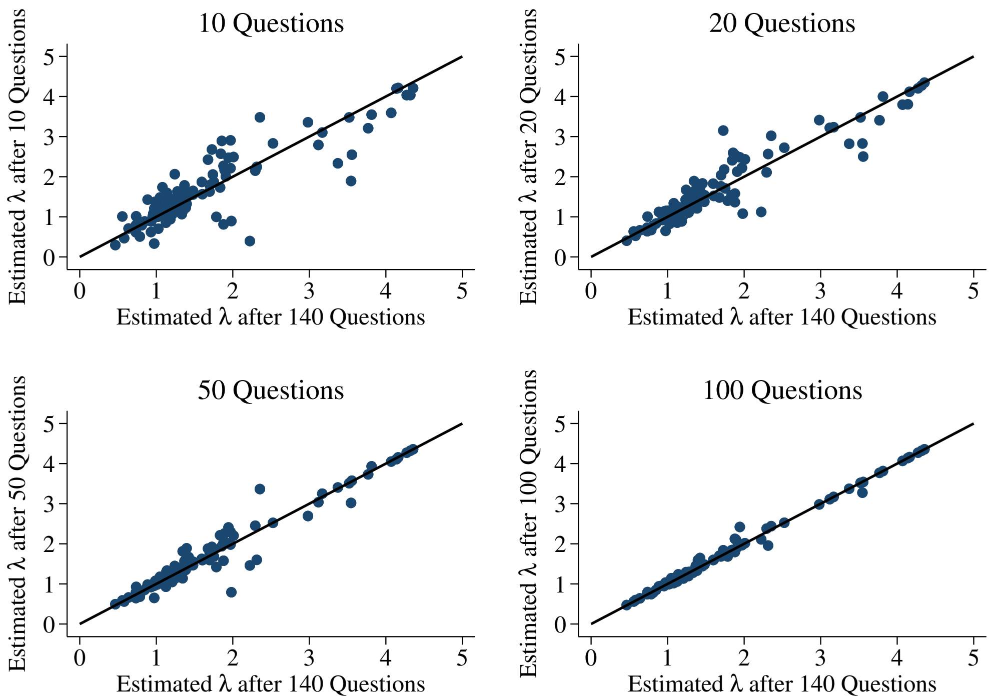


Notes: The figure is based on authors' analysis of data from Sokol-Hessner et al. (2009) and Frydman et al. (2011). Each panel plots the DOSE estimate (using a uniform prior) of the exponent from the utility function (1) against the Bayesian estimate after 140 questions.

C.2 Estimates of Choice Consistency

DOSE provides relatively accurate estimates of the choice consistency parameter as well as risk and loss aversion, as shown in Figure C.4. Compared to the random ordering, the DOSE estimates are closer to and more highly correlated with the final parameter estimate, and are more highly correlated with the final estimate throughout the question sequence. Again,

Figure C.3: Correlations between final estimates of the loss aversion parameter and the estimates after selected rounds.



Notes: The figure is based on authors' analysis of data from Sokol-Hessner et al. (2009) and Frydman et al. (2011). Each panel plots the DOSE estimate (using a uniform prior) of the loss aversion parameter from the utility function (1) against the Bayesian estimate after 140 questions.

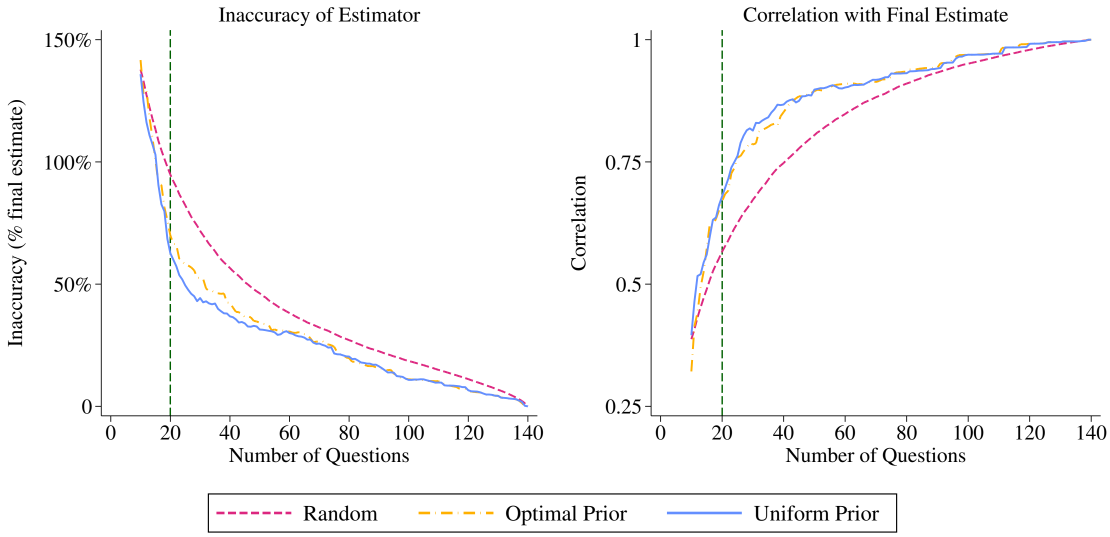
these benefits are similar regardless of whether we use the uniform prior or the "optimal prior" (discussed in the previous subsection).

The choice consistency estimates take longer to converge to the final estimate than the other parameters, likely reflecting the fact that several similar questions must be asked to pin down this parameter. The estimated inaccuracy, for example, after 20 questions is 63%, much higher than the 21–22% for risk and loss aversion (but much lower than the inaccuracy of 94% achieved with the random ordering). Similarly, Figure C.5 shows that after 50 questions the individual-specific estimates are not as closely clustered around the 45 degree line. Again, however, there is no evidence that the accuracy improvements from DOSE are limited to particular values of the choice consistency parameter.

C.3 Maximum Likelihood Estimation

We also attempted to obtain individual parameter estimates using Maximum Likelihood Estimation (MLE), however we were frequently unable to estimate parameters for several

Figure C.4: DOSE elicits accurate estimates for the choice consistency parameter faster than the random ordering.



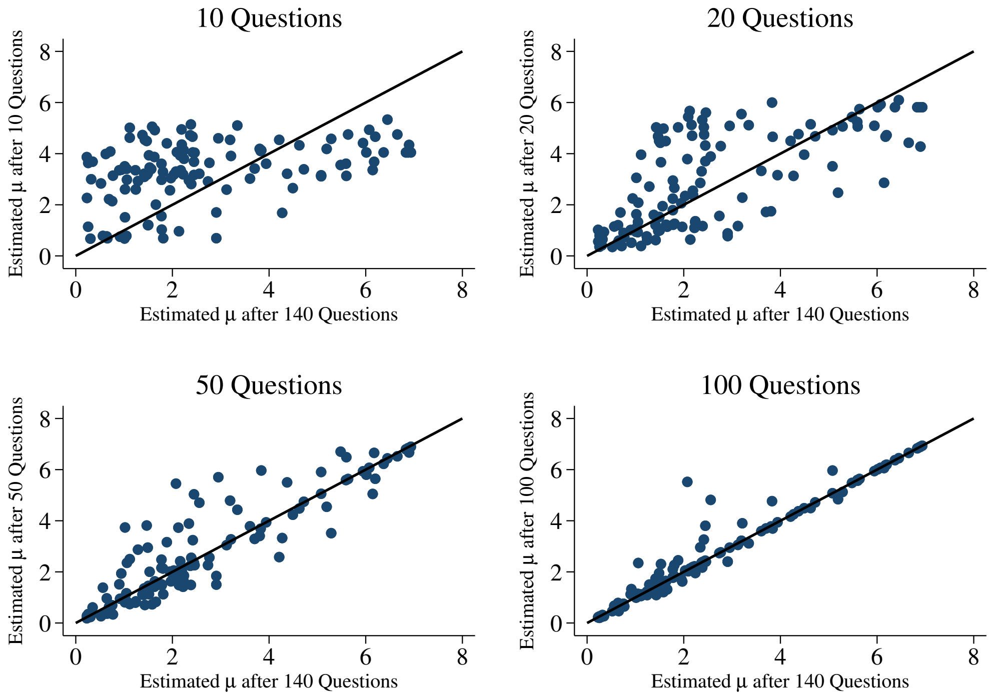
Notes: Based on data from Sokol-Hessner et al. (2009) and Frydman et al. (2011). Left (right) hand panel shows the inaccuracy (correlation with final estimate) of Bayesian estimates obtained after each question, under different orders. "Optimal Prior" and "Uniform Prior" refer to DOSE question selection using corresponding priors. "Random" orders questions randomly, averaging over 100 different random orderings. All parameter values are estimated using a uniform initial prior.

participants.[10] As shown in Figure C.6, when using fewer than 40 questions (using the original order reported in the original datasets), we could not estimate parameter values for one quarter of the sample, and we could not obtain estimates for all participants even when using the full set of 140 questions. This failure is particularly striking given that, for this purpose, we do not exclude any unrealistic values (such as negative parameters) and that, in a final attempt to obtain an estimate, we initiated the search algorithm with the final Bayesian estimate of each individual's parameters. As such these numbers are an overestimate of the proportion of participants for whom meaningful estimates could be recovered in reality; Frydman et al. (2011) in their initial study obtained estimates for only 64 of 83 participants (7 were excluded for other reasons), whereas we report estimates for 82 out of the 90 participants.

Further, the estimates that were obtained by MLE with a small number of questions appear much more inaccurate than those from the Bayesian procedure, as shown by the line plots in Figure C.6. After 140 rounds, the estimates from the different procedures are, as expected, very similar: the correlation between the final MLE and final Bayesian

[10]The MLE procedure was implemented using STATA's modified Newton-Raphson algorithm. Similar results were obtained using alternative algorithms. For each participant estimation was attempted three times (each with up to 16,000 iterations), allowing for alternative initial conditions, different stepping procedures in non-concave regions and relaxing convergence requirements on the gradient vector.

Figure C.5: Correlations between final estimates of the consistency parameter and the estimates after selected rounds.



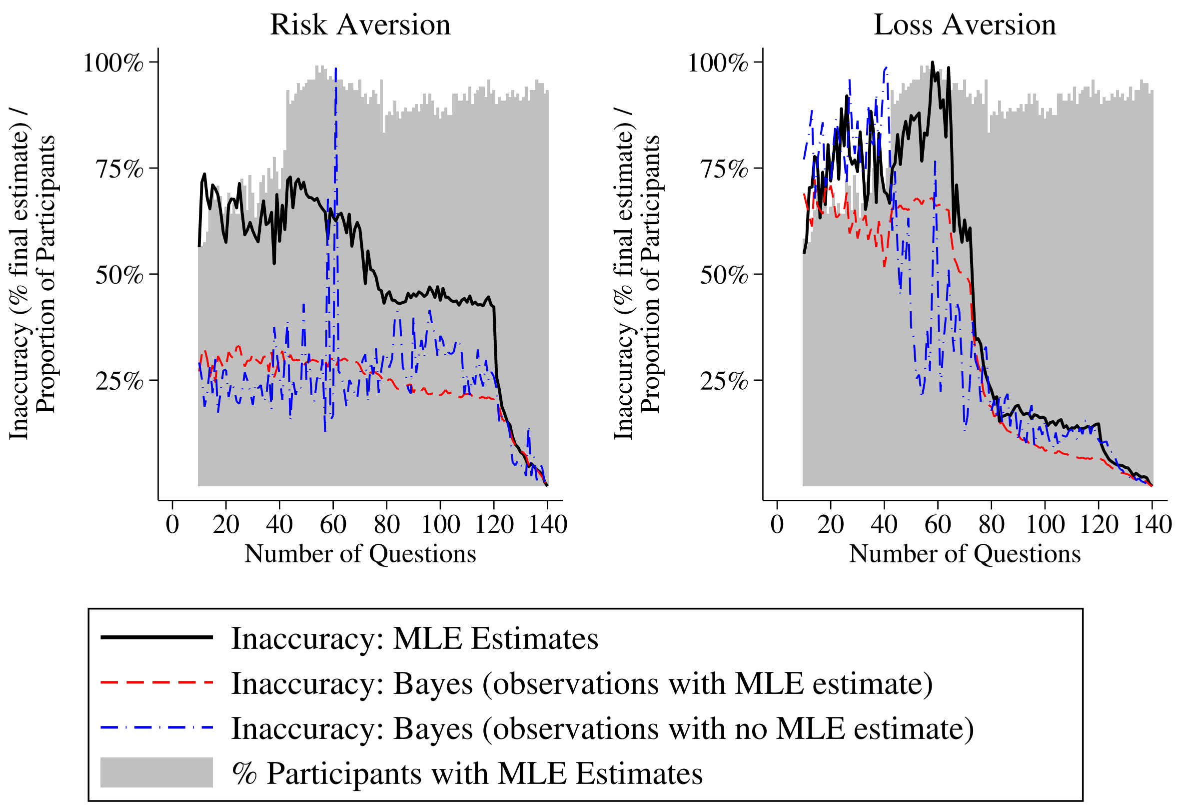
Notes: The figure is based on authors' analysis of data from Sokol-Hessner et al. (2009) and Frydman et al. (2011). Each panel plots the DOSE estimate (using a uniform prior) of the choice consistency parameter in (3) against the Bayesian estimate after 140 questions.

estimates was 0.85 for risk aversion, and 0.95 for loss aversion, while the median distance between the two estimates was less than 2% (of the Bayesian estimate) for both parameters. However, the Bayesian estimates are much closer to these final values after many fewer questions.[11] In addition, the Bayesian estimates are generally more accurate than the MLE estimates that do exist even where no MLE estimate can be obtained at all.[12] Not only can the Bayesian procedure obtain an estimate in those circumstances, those estimates contain valuable information.

[11]To ensure comparability between the two sets of estimates, when calculating the distance from the final estimate we constrain the MLE estimates to the bounds of the prior used for the Bayesian estimates.

[12]The "jerky" nature of the line relating to the inaccuracy when no MLE estimate is available is explained by the fact that—particularly after question 40—few participants do not have MLE estimates, with the precise number varying from round to round. The large spike at round 61, for example, is explained by all but two participants having MLE estimates available.

Figure C.6: With a small number of questions the Bayesian procedure provides more accurate estimates than Maximum Likelihood Estimation.



Notes: The figure is based on authors' analysis of data from Sokol-Hessner et al. (2009) and Frydman et al. (2011). The bars refer to the proportion of participants for whom a parameter estimate could be obtained using Maximum Likelihood Estimation. The lines plot the distance from the estimate obtained after 140 questions after each question round using i) Maximum Likelihood Estimation, ii) Bayesian estimation (where MLE estimates were available), and iii) Bayesian estimation (where MLE estimates were not available).

D Parameter Recovery Exercise

When participants make mistakes DOSE produces estimates that are about twice as accurate as traditional risk and loss aversion elicitation mechanisms. We demonstrate this with a parameter recovery exercise (or Monte Carlo simulation). This is conducted with an entirely simulated dataset that allows us to both know and control the true parameters governing (simulated) participant behavior. We first provide a detailed explanation of the procedure used to simulate DOSE and two other elicitation methods: the double MPL, and Lottery Menu methods. To understand whether our assumptions about the level of noise in the survey are reasonable, we then compare simulated choices to real survey data. The simulation appears to underestimate the level of noise in the survey MPL.

The DOSE estimates are more accurate than other elicitation methods even when the utility function is misspecified. Accurate estimates can still be obtained by using the correct utility function after the fact and, even without re-estimating the results, the DOSE estimates are highly correlated with the true parameter values. Further, DOSE still performs well when estimating a utility function with differential curvature across gains and losses, despite the absence of questions just with losses in our dataset.

D.1 Parameter Recovery Procedure

Simulation Dataset A dataset of 10,000 simulated individuals was generated as follows. First, we estimated the 140 question DOSE procedure on the 120 participants from Sokol-Hessner et al. (2009) and Frydman et al. (2011). We then aggregated the 120 individual posterior distributions to form a joint probability distribution over the three parameters ρ, λ and μ. The 10,000 participants were then drawn from the resulting distribution. In addition, to understand the performance of each elicitation method at different levels of choice consistency, we repeated the simulation assuming each of these simulated individuals had the same μ, fixed at each ventile of the underlying distribution—this analysis is reported in Figure 4.

DOSE Simulation We simulate a 20 question DOSE procedure for each individual, with each binary choice made probabilistically according to the logit probability (3). The possible question space included 760 questions, allowing for gains in \$0.25 increments up to \$10, and losses in \$0.5 increments up to \$10.

Double Multiple Price List (MPL)

We calculate the expected inaccuracy for the double MPL method using two hypothetical MPLs. MPL 1 offers participants a choice between a fixed 50:50 lottery between \$0 and \$10 and a series of fixed amounts. This MPL is used to elicit the estimate of the CRRA coefficient ρ. MPL 2 offers participants a choice between a 50:50 lottery between a loss of \$10 and a gain of \$10 and a series of fixed amounts. This second MPL is used to obtain the estimate of the loss aversion parameter λ. In both MPLs we enforce (in-line with the implementation in the surveys) that individuals could only switch once, and that individuals do not choose dominated options: the left-hand side of MPL (the lottery) is chosen in the first row and the right-hand side (the fixed amount) is chosen in the last row.

The row in which a participant first chooses the fixed amount (the right hand side) in MPL 1 implies a range of certainty equivalents and CRRA coefficients, as shown in Table D.1. We use the certainty equivalent at the midpoint of this range and the associated CRRA coefficient.

Table D.1: Hypothetical MPL 1 used to estimate ρ

Left Hand Choice	Right Hand Choice	CRRA Range	Estimated ρ
50% of \$0, 50% of \$10	\$0	n.a.	n.a.
50% of \$0, 50% of \$10	\$1	$\rho < 0.30$	0.23
⋮	⋮	$0.30 < \rho < 0.43$	0.37
⋮	⋮	$0.43 < \rho < 0.58$	0.50
⋮	⋮	$0.58 < \rho < 0.76$	0.66
⋮	⋮	$0.76 < \rho < 1.00$	1.57
⋮	⋮	$1.00 < \rho < 1.36$	1.16
⋮	⋮	$1.36 < \rho < 1.94$	1.61
⋮	⋮	$1.94 < \rho < 3.11$	1.66
50% of \$0, 50% of \$10	\$9	$3.11 < \rho < 6.58$	1.66
50% of \$0, 50% of \$10	\$10	$6.58 < \rho$	1.66

Notes: "CRRA range" is the implied range of CRRA coefficients implied by the choice of each lottery. "Estimated ρ" is the estimated value of the CRRA coefficient associated with the choice of each lottery used in the calculation of expected inaccuracy. Neither value is defined in the first row because the design does not allow the right hand side to be selected.

Table D.2: Hypothetical MPL 2 used to estimate λ

Left hand choice	Right hand choice
50% of -\$10, 50% of \$10	-\$10
50% of -\$10, 50% of \$10	-\$9
⋮	⋮
⋮	⋮
50% of -\$10, 50% of \$10	\$9
50% of -\$10, 50% of \$10	\$10

Similarly, the row in which a participant first chooses the fixed amount (the right-hand side) in MPL 2 implies a range of certainty equivalents, as shown in Table D.2. We use the certainty equivalent at the midpoint of this range and use the estimated CRRA coefficient $\hat{\rho}$ estimated in MPL 1 to obtain the estimated loss aversion parameter, $\hat{\lambda}$. For comparability with the DOSE estimates, we truncate the range of $\hat{\lambda}$ and $\hat{\rho}$ to match the range of the prior used in the DOSE procedure.

The procedure for simulating behavior on these two MPLs was as follows. For each row

r, the probability that a simulated individual defined by the parameter vector (ρ, λ, μ) first chooses the right hand side of the MPL in row r is calculated. This probability is defined by the logit probability (see (3)) comparing the lottery to the fixed amount offered in row r. To translate these binary choices into a probability distribution over the set of rows in the MPL we assume that individuals work either sequentially down or up an MPL, each with 50% probability. Suppose they work down the MPL. Then they first consider the choice between the lottery and the fixed amount in the first row in which they can choose the fixed amount (row 2 in our implementation). If they choose the fixed amount, they will always prefer the fixed amount lower in the MPL: thus this row is the "switching row". If, on the other hand, they prefer the lottery then they will move to the next row and consider the next binary choice. Alternatively, individuals may choose to work up the MPL by first considering the bottom row of the MPL, then the second-bottom, etc.

Now consider a MPL with $\mathcal{R}$ rows in which an individual can switch. Define the probability that the lottery is chosen in row r by individual i as q_r^i. This probability is defined by ρ, μ and, when losses are involved, λ. For simplicity we suppress the i indices. Define the probability row r is the switching row working down the MPL as p_r^D, and working up the MPL as p_r^U. Then these probabilities are given by:

$$\begin{aligned} p_r^D &= (1-q_r)\prod_{s=1}^{r-1} q_s \quad \text{and} \\ p_r^U &= (q_{r-1})\prod_{s=r}^{\mathcal{R}}(1-q_s) \end{aligned}$$

The expected inaccuracy for any parameter θ is then given by:

$$E[|\hat{\theta}_r - \theta|] = \sum_{r=1}^{\mathcal{R}} \left(0.5p_r^D + 0.5p_r^U\right) |\hat{\theta}_r - \theta|$$

where $\hat{\theta}_r$ is the estimated parameter associated with switching in row r. As discussed above, for ρ this is implied by the midpoint of the certainty equivalents defined by the switching row. For λ the value is defined both by the midpoint of the certainty equivalent and the estimated $\hat{\rho}$ from MPL 1.

Risky Project In the risky project, developed by Gneezy and Potters (1997), participants are given an endowment of \$2 and offered the chance to invest any amount they choose in a risky project. Any money they choose not to invest they keep. With 40% probability the project is successful and the participant receives three times the amount they invested (plus any money they chose not to invest). If the project fails, they lose their investment.

The procedure for the simulation was as follows. A participant starts by considering a choice between investing \$0 and investing the full \$2, with the probability of each choice defined by the logit function (3). They then compare the winner of this initial choice (\$0 or \$2) to investing \$1 (half the possible investment). If they choose \$0, they invest \$0. If they choose \$2, they then make sequential comparisons between lower possible investment amounts, using increments of \$0.20. So they first compare \$2 to \$1.80—if they choose \$2, then they stop and invest \$2. If they choose \$1.80, they compare \$1.80 to \$1.60 etc, and so on in increments of \$0.20. Similarly, if they choose \$1 they then sequentially compare lower amounts, starting with \$1 vs \$0.20. The use of increments of \$0.20 captures the fact that, as we will see in the following sub section, the distribution of participants' investments in our survey data appears discrete rather than continuous. The results are similar using alternative increments.

This procedure produces a probability distribution for each individual i, over eleven possible investment amounts m ranging from \$0 and \$2 in increments of \$0.20—that is, $m \in M = \{0, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2\}$—with the probability of investing m given by p_m. The amount of points invested (m) defines an estimated $\hat{\rho}_m$. To ensure comparability with the DOSE estimates, we then truncate the estimated parameters to the range defined by the Sokol-Hessner-Frydman distribution. The estimated inaccuracy is then given by:

$$E[|\hat{\rho} - \rho|] = \sum_{m}^{M} p_m \times |\hat{\rho}_m - \rho|$$

Lottery Menu In the lottery menu procedure, developed by Eckel and Grossman (2002), participants are offered a choice between multiple lotteries over gains. We calculate the expected measurement error for the menu of six 50:50 lotteries presented in Table D.3. This implementation is based on the menu used by Dave et al. (2010), adjusted so that the largest prize is \$10 (for comparability with the other elicitation procedures). The first lottery is a safe option (it has zero variance), while the subsequent lotteries increase in both expected value and variance.

The choice of lottery implies a range of possible CRRA coefficients, as shown in the penultimate column of Table D.3. For lotteries 2–5 we estimate the estimated CRRA coefficient $\hat{\rho}$ as the midpoint of this range. Since the midpoint is undefined for lotteries 1 and 6, for these lotteries we use the end-point of the range. To ensure comparability with the DOSE estimates, we then truncate the estimated parameters to the range defined by the

Table D.3: Choices in Simulation of Lottery Menu Procedure

	Low Prize	High Prize	CRRA Range	Estimated ρ
Lottery 1	4.00	4.00	$\rho < -2.46$	0.20
Lottery 2	3.43	5.14	$-2.46 < \rho < -0.16$	0.20
Lottery 3	2.86	6.29	$-0.16 < \rho < 0.29$	0.20
Lottery 4	2.29	7.43	$0.29 < \rho < 0.50$	0.40
Lottery 5	1.71	8.57	$0.50 < \rho < 1.00$	0.75
Lottery 6	0.29	10.00	$1 < \rho$	1.00

Notes: Lottery menu choices taken from Dave et al. (2010), adjusted so that maximum prize is \$10. "CRRA range" is the implied range of CRRA coeffcients implied by the choice of each lottery. "Estimated ρ" is the estimated value of the CRRA coefficient associated with the choice of each lottery used in the calculation of expected inaccuracy.

Sokol-Hessner-Frydman distribution.

The procedure for the simulation was as follows. Consider a menu over a set of lotteries $l_1, l_2, ..., l_{\mathcal{L}}$. We define a probability distribution over the set of lotteries by assuming that individuals make a series of binary choices in which they compare the set of lotteries in order. That is, they first compare lottery 1 with lottery 2, making a choice according to the logit probability. They then compare the winner of that choice with lottery 3, and then the winner of the latter choice with lottery 4. The procedure is repeated until lottery $\mathcal{L}$.

We define a probability distribution over the full lottery menu for each participant i as follows. For two lotteries l, k let $q^i_{l,k}$ be the probability that i chooses l when faced with a binary choice between l and k. This probability is defined by the logit function (3), and as such depends on the participant's value of ρ and μ; for simplicity we do not display these parameters or the i index in the following. Define the probability that lottery l is chosen after L choices as p^L_l. Then $p^1_1 = q_{1,2}$ and for all other l, L:

$$p_l^L = \sum_{k=1}^{l-1} q_{l,k} p_k^{l-1} \times \prod_{m=l+1}^{L} q_{l,m}$$

The probability distribution over the choice from the set of lotteries is then $\{p_1^{\mathcal{L}}, p_2^{\mathcal{L}}, ..., p_{\mathcal{L}}^{\mathcal{L}}\}$. Defining $\hat{\rho}_l$ as the estimated CRRA coefficient associated with a choice of lottery l, the expected inaccuracy is given by:

$$E[|\hat{\rho} - \rho|] = \sum_{l=1}^{\mathcal{L}} p_l^{\mathcal{L}} \times |\hat{\rho}_l - \rho|$$

We also implemented an alternative simulation procedure for the Lottery Menu. Under this alternative, choice occurred according to to a multinomial logit probability distribution. That is, for each possible choice $k = 1, ..., 6$:

$$Prob(Choice = k) = \frac{exp(EU_k)^{\mu}}{\sum_{l=1}^{6} exp(EU_l)^{\mu}}$$

where EU_k is the expected utility of lottery k.

The simulation results for the Lottery Menu procedure were similar to that for the risky project. Using our main simulation (presented in Table 1), the average inaccuracy was 35% (or 42% using the multinomial logit function for choice), and the correlation with the true parameter was 0.29 (0.37). For the lowest consistency ventile (as in Figure 4) the average inaccuracy was 41% (55%) and in the highest ventile 27% (29%).

D.2 Misspecification of the Utility Function

The DOSE estimates are robust to misspecification of the utility function. We run DOSE on the same 10,000 simulation subjects—each of whom has CRRA utility—but assuming a CARA utility function in the question selection procedure. We then compare the correlation between the risk aversion and loss aversion parameters under the different procedures, and demonstrate how the data collected can be re-estimated to elicit accurate CRRA utility parameters.[13]

Misspecifying the utility function does not lead to a loss of accuracy, as shown in Table D.4. For risk aversion, we can recover the same estimates by re-estimating the correct utility function after the data collection process. For loss aversion, very similar estimates are obtained even when the CARA function is incorrectly used.

Further, the assumptions over parametric form are unlikely to be critical if researchers are interested in identifying correlations rather than the level of the risk and loss aversion estimates. Even without re-estimating, the Spearman correlation between the estimated CARA parameters and the true (CRRA) parameter values is very high—and notably higher than the correlations for either the MPL or the risky project procedures reported in Table 1.

[13]Specifically we run DOSE on the simulation dataset assuming the following exponential (CARA) utility function:

$$u(x, \gamma_i, \lambda_i) = \begin{cases} \frac{1-e^{-\gamma_i x}}{\gamma_i} & \text{for } x \geq 0 \\ \lambda_i \left(\frac{e^{\gamma_i x}-1}{\gamma_i} \right) & \text{for } x < 0 \end{cases} \tag{5}$$

where λ represents loss aversion and γ captures risk aversion.

Table D.4: DOSE estimates are robust to utility function misspecification.

	Average inaccuracy		Correlation with true value	
	10 question	20 question	10 question	20 question
Risk Aversion				
CRRA (Not misspecified)	22%	16%	0.65	0.78
CARA (Misspecified)	n.a.	n.a.	0.59	0.74
CARA re-estimated as CRRA	21%	17%	0.65	0.76
Loss Aversion				
CRRA (Not misspecified)	22%	16%	0.85	0.90
CARA (Misspecified)	24%	18%	0.83	0.90
CARA re-estimated as CRRA	23%	17%	0.84	0.90

Notes: Inaccuracy is defined as the absolute distance from the true parameter value displayed as a percentage of the true value. "Correlation with true value" displays the Spearman correlation coefficient between the true parameter and the estimated parameters.

E Robustness Checks

This section presents extended survey results and robustness tests. In the first subsection, we demonstrate that our main results are robust to misspecification. We then examine the DOSE choice data, and show that the ranking of individuals' preferences is very similar when re-estimating with alternative utility or error specifications. The third sub-section then shows that our main conclusions are robust when using these alternative estimates. The fourth subsection includes additional regression results, showing that the correlation between cognitive ability and economic preferences is robust to alternative definitions of control variables, and also robust to removing participants least likely to be paying attention. The final subsection (E.5) shows that DOSE choice consistency measure can identify inattention (Section 5.3) even when restricting the sample to those answering quickly—a standard way of identifying inattention in the survey.

E.1 Additional Tests of Fatigue and Inattention

Neither survey fatigue or inattention appear to explain our results. The pattern of economic preferences does not change according to the position of the DOSE module, suggesting that participants did not become fatigued later in the survey. Nor do participants appear to choose more randomly later in the DOSE module, as would be expected if they get bored of the questions. There is also little evidence that participants get tired and click through the survey quickly without paying attention: there is little correlation between response time and our three economic preference parameters. Further, we observe the same relationships between economic preferences and individual characteristics when excluding the fastest participants. Thus, boredom causing unusually fast choice, which then leads to our results does not seem a plausible explanation for our conclusions.

It does not appear that survey fatigue affected our estimate of any of the four DOSE parameters, as shown in Figure E.1. The position of the two DOSE modules was randomized across participants, and could appear in the third, fourth, fifth or sixth position in the survey (the two modules always appeared together). The distribution of parameters is similar regardless of the position in both waves of the survey. Thus it does not appear that participants' behavior changed when they took DOSE further into the survey, suggesting that fatigue is not a major factor in our results.

We can also look for signs of fatigue within the DOSE module itself by observing that if participants start choosing randomly then the Bayesian prior will be "surprised" more often. That is, at each point in the question selection process, DOSE's prior has an alternative that it thinks the participant will choose. When the participant chooses an option the procedure thinks should only be a 30% chance, we encode that as $50 - 30 = 20$ percentage points surprising. If, on the other hand, the participant chooses an option the procedure things is 70%, then we encode this as $50 - 70 = -20\%$ surprising. If participants are getting fatigued, we would expect them to start choosing randomly, that is, make more surprising choices (on average) later in the module.

As can be seen in Figure E.2, there is no trend of increasing surprise towards the end of either module. Surprise is, on average, slightly lower at the end of the time preference module, suggesting that if anything, participants are making more consistent choices towards the end of the procedure than at the beginning. Notably, the procedure is particularly surprised in question 4, when participants are presented with a question involving losses for the first time—reflecting the fact that, as we detail in depth in Chapman et al. (forthcoming), we observe much lower levels of loss aversion in our general population sample than in prior laboratory studies. Since our prior was based on those laboratory studies, the procedure "expected" a much higher degree of loss aversion than was actually the case.

While there is little evidence of survey fatigue, it is possible that some participants were inattentive throughout the survey. We examine this possibility next by reanalyzing our data while leaving out those who were most likely to have given up and rushed through the survey. One might expect that bored or inattentive participants would just click through screens quickly but, as shown in Figure E.3, our results are largely unchanged when removing the fastest responses. In this figure, we first look at the slowest 80% of participants, then the slowest 60%, and so on. The distributions overlap almost entirely. The parameter distributions are similar, suggesting that neither confusion nor inattentiveness, nor giving up, is likely to explain many of the choices we see.

E.2 DOSE Choice Data and Model (Mis-)Specification

Our results are underpinned by a clear set of choices, and hence re-estimating with alternative functional forms for utility or choice does not significantly affect our results. Our estimates of each of risk and loss aversion are highly correlated when re-estimating using CARA utility, allowing differential risk aversion over gains and losses, or when implementing either a probit error specification or the random parameter model suggested by Apesteguia and Ballester (2018). Consequently, as we show in the following subsection, alternative specifications make little difference to correlations between these characteristics and cognitive ability.

In Table E.1 we classify participants according to their estimated parameter values—for instance, a participant is "loss averse, risk averse" if they have both $\lambda > 1$ and $\rho < 1$—and examine how the frequency of lotteries accepted varies according to the expected value (relative to a sure amount) and whether the lottery involved a loss. The pattern of behavior is as would be expected. Loss tolerant participants nearly always choose lotteries with losses, and risk loving participants nearly always choose lotteries over gains. Loss averse and risk averse participants, in contrast, are much less likely to accept such lotteries.

In fact, our DOSE parameter estimates are highly correlated with non-parametric preference measures constructed from the DOSE data. The Spearman correlation between the DOSE risk aversion parameter and the percentage of lotteries accepted offering only potential gains is 0.92. The DOSE loss aversion parameter is correlated 0.90 with the percentage of mixed lotteries accepted. And the DOSE patience parameter is correlated 0.84 with the share of longer-dated payments accepted in the time preference DOSE module.

Given these facts, it is not surprising that our estimates of participants' preferences are similar when we re-estimate the DOSE parameters with alternative parametric assumptions. We re-estimate risk and loss aversion using first a CARA utility function, and then a CRRA utility function in which we allow for differential utility curvature between gains and losses.

Table E.1: DOSE classification reflects a clear pattern of choices.

	% Lotteries Accepted			
	Lotteries with Losses		Lotteries with Only Gains	
	EV≤Sure	EV>Sure	EV≤Sure	EV>Sure
Classification by DOSE				
Loss averse, risk averse	14%	46%	8%	41%
Loss averse, risk loving	2%	51%	53%	98%
Loss tolerant, risk averse	82%	99%	5%	41%
Loss tolerant, risk loving	82%	100%	77%	96%

Notes: The table displays the unweighted percentage of lotteries accepted, categorizing participants according to their estimated DOSE parameters. "EV"=Expected Value of lottery and "Sure"= the sure amount offered in each lottery.

The Spearman correlation between the estimates of risk aversion (over gains) from these specifications and our main specification is 0.98 in both cases. For loss aversion it is 0.98 when assuming CARA utility, and 0.76 when allowing for differential curvature—lower, as we might expect given that part of behavior over losses is now captured by the additional parameter, but still highly correlated with our main estimates. Consequently, as we will see below, our estimated correlations with cognitive ability are very similar when using these alternative utility functions to construct parameter estimates.

To understand whether our logit error function plays an important role in our results, we re-estimate the DOSE parameter estimates with a probit error specification (Figure E.4) and then a random parameter model (Figure E.5). As pointed out by Apesteguia and Ballester (2018), the logit error function is potentially problematic in the estimation of risk aversion because the relationship between the level of risk aversion and the probability of accepting a lottery is non-monotonic—potentially creating issues for identification of the risk aversion parameter. Those non-monotonicities do not affect loss aversion directly because the difference between the utility of a lottery with a loss and a sure amount is monotonically decreasing in loss aversion, meaning that higher loss tolerance is associated with an increasing probability of accepting a lottery (for a given level of risk aversion). However, they could impact our loss aversion estimates indirectly through the risk aversion parameter. As such, we check whether the error specification we use has a major impact on estimated risk preferences using both a probit error function and the random parameter model suggested by Apesteguia and Ballester (2018).

First, Figure E.4 shows our results for risk and loss aversion using a probit (rather than logit) error function. That is, we assume that the probability that the lottery is chosen follows a standard normal distribution over the utility difference between the lottery and sure amount. As we can see, the estimates are very highly correlated.

Second, we implement the RPM, by estimating risk aversion and then, in a second stage, estimated loss aversion. We choose this two stage approach as, first, implementing a RPM in a two parameter model is not straightforward (Apesteguia and Ballester (2018) examine only a one parameter model) and, second, it is only risk aversion that is directly affected by the non-monotonicities (as discussed in the previous paragraphs). The risk aversion parameter is largely identified using questions where the lottery contains only gains. As such we used a RPM on those questions to identify the risk aversion parameter. We then fix individual-level risk aversion parameters using the RPM estimates, and implement the logit specification to estimate loss aversion using the remaining questions—those that ask participants to choose between a lottery with both a loss and a gain and a sure amount of zero.

Our results are also similar when implementing the RPM, as shown in Figure E.5. The main difference is that, as shown in the left-hand panel, the RPM risk aversion estimates are (almost) uniformly lower numerically than when using the logit specification—but the estimates are extremely highly correlated (0.98) with those from the logit specification. These numerical differences in the risk aversion parameter do not, however, significantly impact our loss aversion estimates, as shown in the right hand panel. The correlation with our main estimates is again very high (0.99).

E.3 Robustness to Possible Misspecification

This subsection demonstrates that our findings regarding the the relationship between economic preferences and cognitive ability are robust to possible misspecification of the utility function or the error specification used to obtain the DOSE parameter estimates. As discussed in the previous section, DOSE estimates are underpinned by a clear pattern of choices, and hence alternative functional forms do not significantly change the ranking of individuals' preferences. Consequently, our findings regarding cognitive ability are largely unchanged when re-estimating the results using alternative parametric forms.

There are clear differences in the pattern of choices according to level of cognitive ability. Consistent with the correlations presented in Table 2, high cognitive ability participants are more likely to accept lotteries offering only gains, less likely to accept mixed lotteries (i.e., those over both gains and losses), and more likely to accept later-dated payments. In particular, participants in the top tercile of cognitive ability accepted 46% of lotteries over only gains, compared to 40% for those in the middle tercile, and 46% for those in the bottom tercile. The share of mixed lotteries accepted was 58%, 68%, and 70% for those in the top, middle, and bottom tercile respectively. The share of later-dated payments accepted was 78% for the top tercile, 72% for the middle tercile, and 67% for the bottom tercile.

Figure E.6 shows that the correlation between cognitive ability and each of risk and loss aversion is similar when allowing for the alternative functional forms for utility or the error process discussed in the previous subsection. Specifically, the figure displays the coefficients from estimating the regressions in Table 2 with no controls (that is, a univariate correlation) and with the full set of controls, for both risk aversion and loss aversion. We can see that the estimated coefficients are very similar under the different specifications, both in magnitude and statistical significance. The stability of these estimates are perhaps not surprising, given the pattern of choices discussed in the previous paragraph. However, in other settings, the ability to re-estimate results ex post, if the parametric form used for question selection turns out to be inappropriate, may be a valuable tool for researchers.

Table E.2 replicates the regressions in Table 2 allowing for present bias. The correlations between each characteristic and discounting are similar regardless of whether we allow for present bias. Present bias is, similarly to our other preference estimates, strongly correlated with cognitive ability—those with higher cognitive ability are less present biased. However, there are some differences between discounting and present bias when examining the relationship with other variables. In contrast to discounting, there is no evidence of a relationship between present bias and education after controlling for cognitive ability but there is a correlation with income. Further, age appears strongly related to discounting, but not to present bias.

E.4 Robustness of Correlations with Economic Preferences

In this subsection we present extended versions of the correlation tables in Section 5, and robustness tests of the relationship between cognitive ability and economic preferences. The relationships between economic preferences and cognitive ability are robust to controlling for other individual characteristics, including both income and education. Indeed, most of the correlation between risk and loss aversion and education is explained by differences in cognitive ability.

In Table E.3 we compare the correlations when using the DOSE measure of risk aversion (column 1) and time preference (column 6) with the other risk and time measures in our survey. For risk aversion these alternative measures included two MPL modules, one relating to Willingness-to-Pay for a lottery (which we use in Section 5.3), and one relating to Willingness-to-Accept, as well as a risky project measure (Gneezy and Potters, 1997).

Table E.2: Cognitive ability is strongly correlated with discounting and present bias.

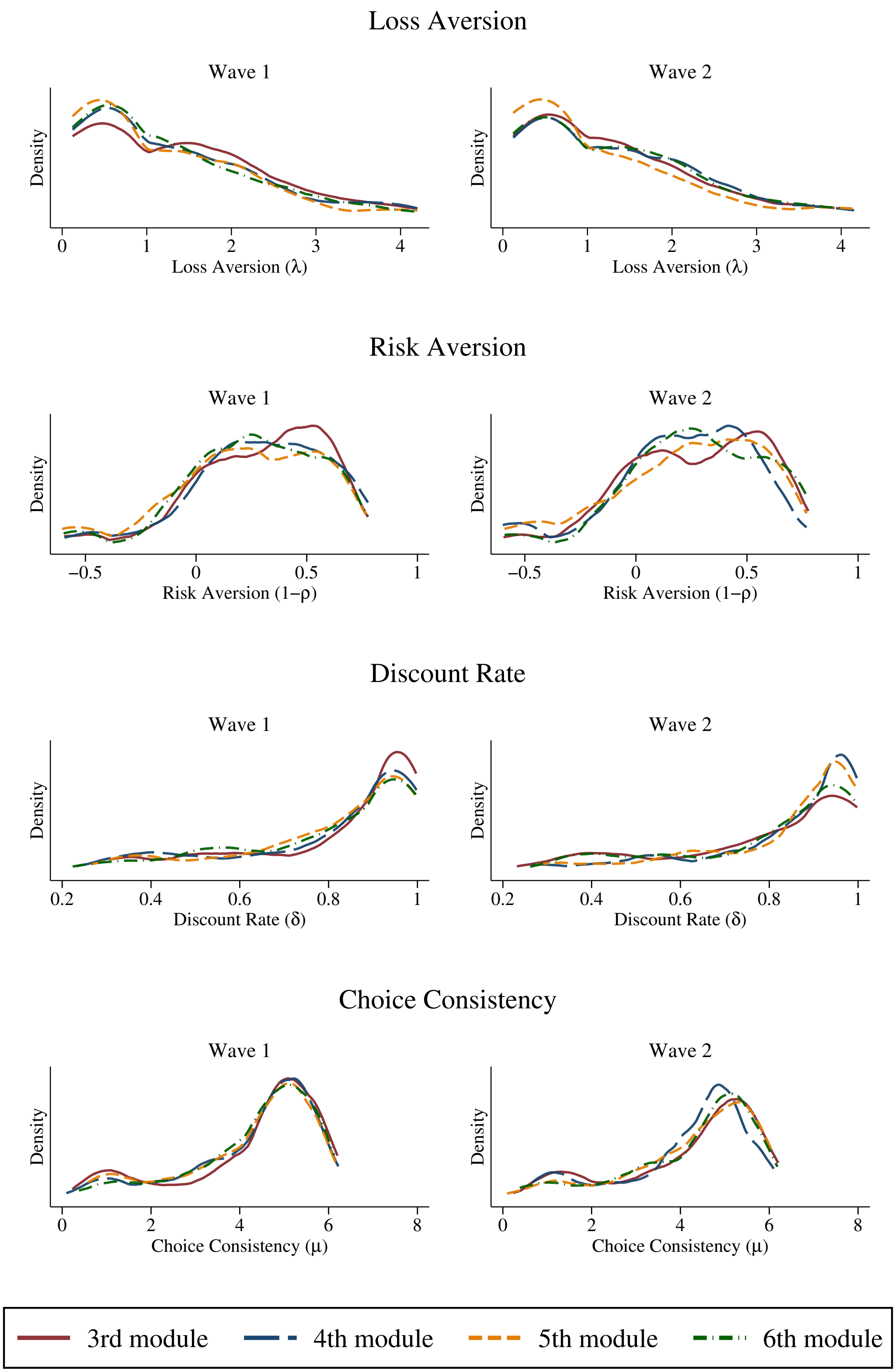
	No Present Bias		With Present Bias			
	Patience (δ)		Patience (δ)		Present Bias (β)	
	Univariate Correlation	Multivariate Regression	Univariate Correlation	Multivariate Regression	Univariate Correlation	Multivariate Regression
Cognitive Ability	0.18***	0.18***	0.17***	0.15***	0.17***	0.14***
	(0.029)	(0.028)	(0.030)	(0.029)	(0.027)	(0.029)
Education	0.17***	0.11***	0.16***	0.11***	0.11***	0.03
	(0.037)	(0.038)	(0.036)	(0.038)	(0.032)	(0.033)
Income	0.09***	-0.02	0.10***	-0.00	0.15***	0.10**
	(0.036)	(0.040)	(0.035)	(0.040)	(0.039)	(0.047)
Male	-0.02	-0.05*	-0.01	-0.04	0.05	0.01
	(0.035)	(0.033)	(0.035)	(0.032)	(0.033)	(0.032)
Age	0.18***	0.18***	0.15***	0.15***	0.01	0.00
	(0.036)	(0.036)	(0.035)	(0.035)	(0.035)	(0.035)
Married	0.09***	0.03	0.08***	0.03	0.05	0.00
	(0.035)	(0.038)	(0.035)	(0.038)	(0.033)	(0.039)
N	2,000	2,000	2,000	2,000	2,000	2,000

Notes: ***, **, * denote statistical significance at the 1%, 5%, and 10% level. Robust standard errors, in parentheses, come from a standardized regression. The first, third, and fifth columns report univariate correlations. The second, fourth, and sixth report the coefficient from a multivariate regression.

For time preferences, as discussed in Section 2, we included two MPLs as well as the DOSE module.

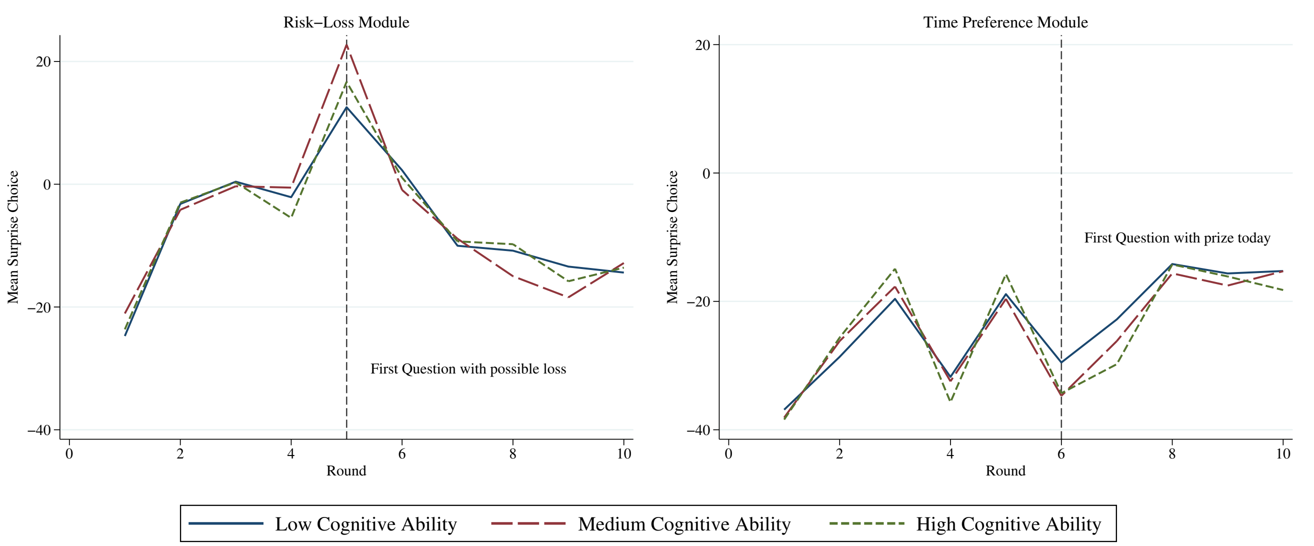
The pattern of correlations is much stronger when using the DOSE measure than either MPL measure. As discussed in Section 5.3, the weak correlations with the MPL (WTP) measure are consistent with attenuation bias due to higher measurement error in the MPL. The weak pattern of correlations with the MPL (WTA) measure could also be explained by attenuation bias or could result from the WTA measure capturing a different dimension of risk preferences to the other risk measures in our survey (see Chapman et al., 2023). The risky project measure, which may suffer from less attenuation bias than the MPLs due to its simplicity, identifies a similar pattern of correlations to the DOSE risk aversion measure. The correlations between the risky project measure and individual characteristics consistently have the same sign, degree of statistical significance and magnitude as those with the DOSE estimates. The main exception are the correlations with cognitive ability, where DOSE identifies much stronger correlations than the project measure.

Figure E.1: The parameter distributions are similar regardless of the position in the survey.



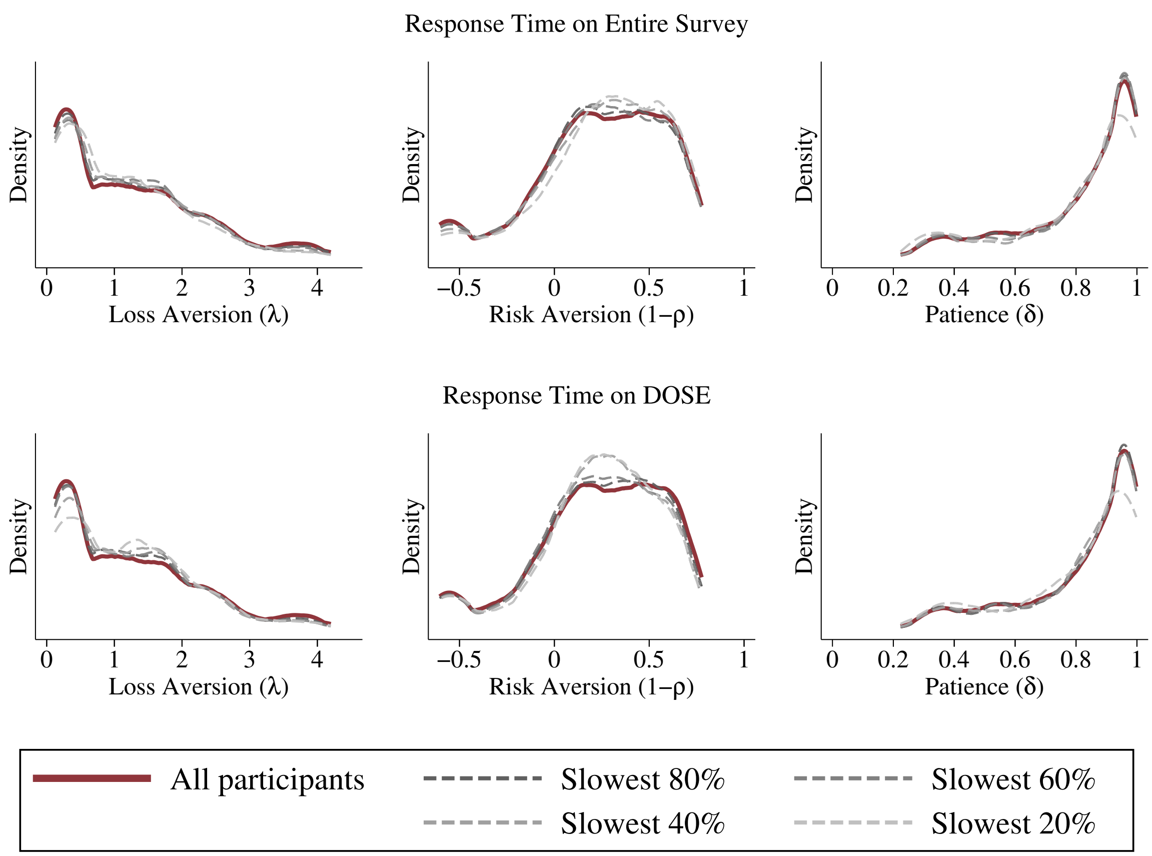
Notes: Figure displays the kernel density of each parameter using an Epanechnikov kernel, according to the position of the DOSE module in the survey.

Figure E.2: No evidence of fatigue within the DOSE module.



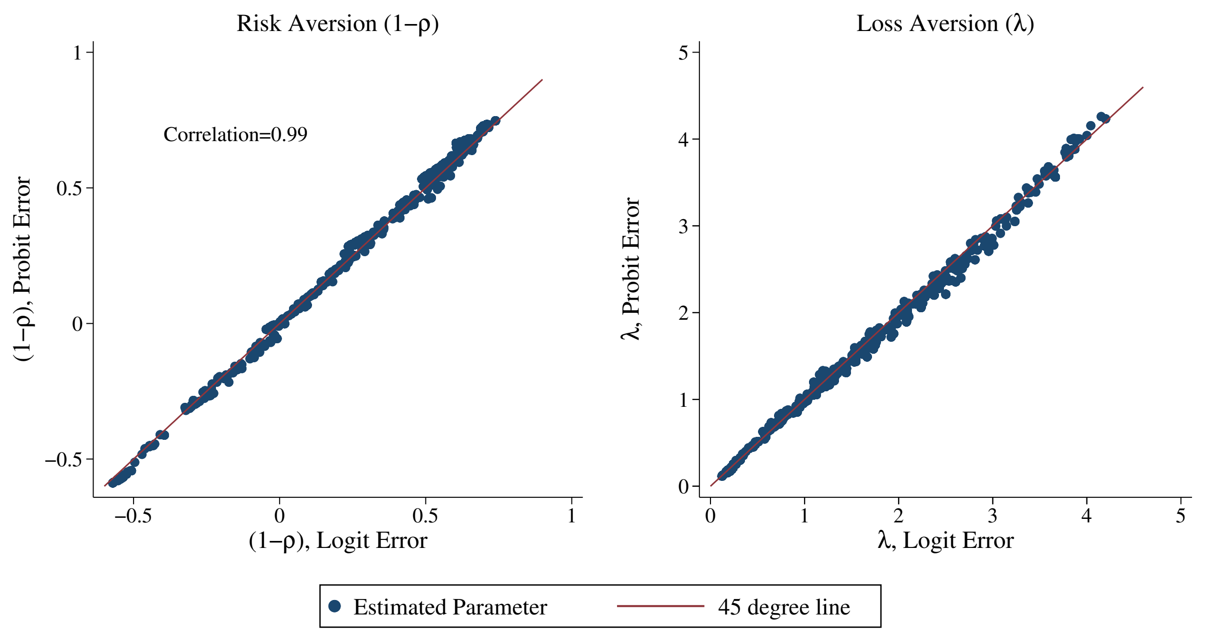
Notes: The figure plots the mean "surprise" of individuals' choice in each round, where surprise is calculated using the DOSE priors before each question. Questions with losses were allowed only from round 5 onwards, leading to considerable updating and a high level of surprise. In the time preference module, questions with payments today were introduced only from round 6 onwards.

Figure E.3: Distributions are similar when removing participants with short response times.



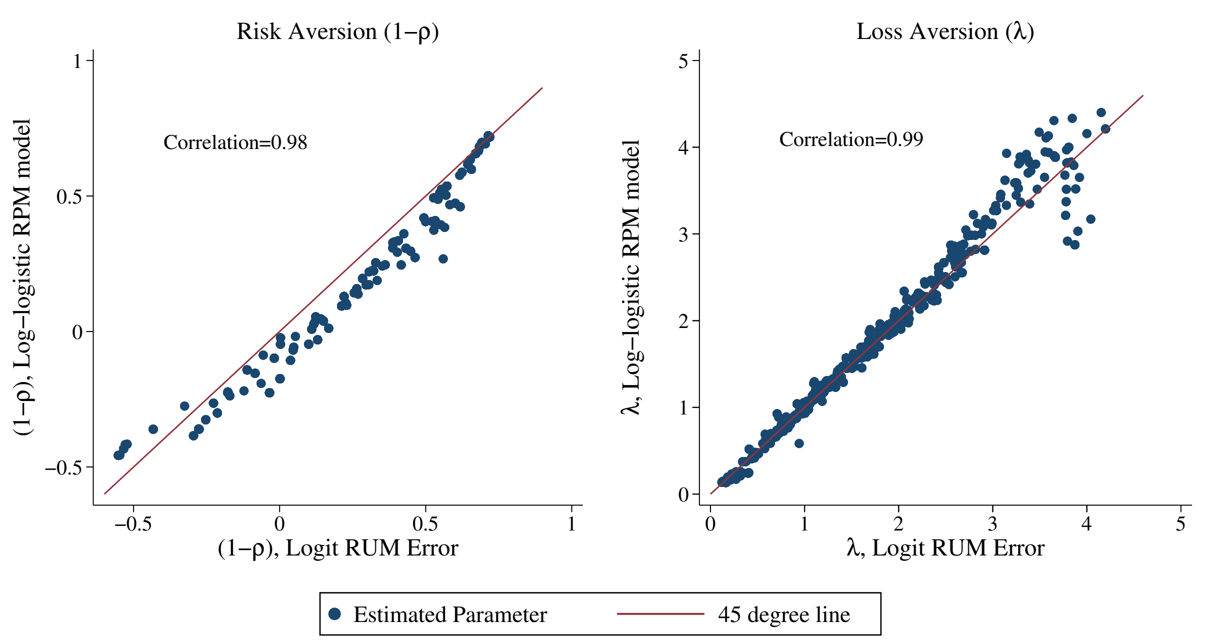
Notes: Plotted using Epanechnikov kernel, with bandwidth chosen by rule-of-thumb estimator.

Figure E.4: Using the probit error specification leads to very similar parameter estimates.



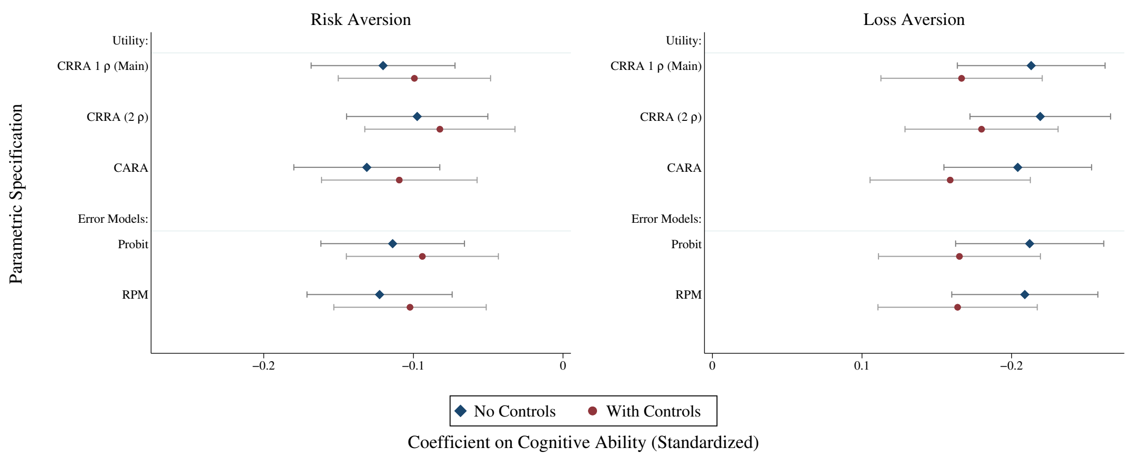
Notes: The figure displays the distribution of the loss and risk aversion estimates under the logit error specification used in the main estimates (x-axis) and a probit error specification (y-axis).

Figure E.5: Parameter estimates are similar using a random parameter model.



Notes: The left-hand panel of the figure displays the DOSE estimates of risk aversion using only lotteries over gains under the Random Parameter Model (RPM; y-axis) and Random Utility Model (RUM; x-axis). The right-hand panel displays estimates of loss aversion from the same two models.

Figure E.6: Correlations between risk preferences and cognitive ability are robust to mis-specification of utility function or error process.



Notes: The figure displays the correlations between cognitive ability and each of risk and loss aversion, using preference parameters from the risk-loss DOSE module. "No controls" represents univariate correlations, while "with controls" includes the set of control variables in Table 2. Bars represent 90% confidence intervals.

Table E.3: Comparison of correlations between different risk and time measures and individual characteristics

	Risk Aversion					Patience		
	DOSE	MPL C.E. (WTP)	MPL C.E. (WTA)	Risky Project	Self-Assessment	DOSE (δ)	MPL (δ)	Self-Assessment
Cognitive Ability	-0.21***	-0.04	0.01	-0.07***	0.01	0.18***	0.19***	-0.10***
	(0.028)	(0.028)	(0.028)	(0.029)	(0.030)	(0.029)	(0.027)	(0.030)
IQ	-0.18***	-0.05	0.00	-0.07**	0.00	0.14***	0.17***	-0.09***
	(0.029)	(0.028)	(0.029)	(0.030)	(0.031)	(0.032)	(0.029)	(0.031)
CRT	-0.18***	-0.02	0.02	-0.05*	0.02	0.18***	0.15***	-0.08***
	(0.029)	(0.027)	(0.025)	(0.029)	(0.034)	(0.036)	(0.025)	(0.031)
Income (Log)	-0.11***	-0.05	0.01	-0.12***	-0.03	0.10***	0.08***	0.03
	(0.035)	(0.032)	(0.032)	(0.035)	(0.035)	(0.035)	(0.033)	(0.032)
Education	-0.10***	-0.02	0.03	-0.09***	-0.06*	0.17***	0.11***	-0.01
	(0.033)	(0.034)	(0.028)	(0.032)	(0.035)	(0.037)	(0.033)	(0.033)
Male	-0.10***	-0.06**	0.01	-0.12***	-0.12***	-0.02	0.03	-0.03
	(0.032)	(0.030)	(0.030)	(0.032)	(0.033)	(0.035)	(0.034)	(0.032)
Age	0.01	0.05	-0.04	-0.02	0.13***	0.18***	0.15***	0.12***
	(0.032)	(0.031)	(0.028)	(0.032)	(0.035)	(0.036)	(0.035)	(0.033)
Stock Investor	-0.11***	-0.06**	-0.00	-0.12***	-0.04	0.10***	0.09***	0.01
	(0.029)	(0.029)	(0.026)	(0.030)	(0.030)	(0.031)	(0.030)	(0.030)
Own Home	-0.07**	0.02	-0.04	-0.08***	0.03	0.13***	0.11***	0.10***
	(0.032)	(0.030)	(0.030)	(0.032)	(0.034)	(0.035)	(0.034)	(0.033)
Married	-0.01	0.02	0.01	-0.02	0.02	0.09***	0.05	0.11***
	(0.032)	(0.030)	(0.030)	(0.032)	(0.034)	(0.035)	(0.034)	(0.032)
N	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000

Notes: ***, **, * denote statistical significance at the 1%, 5%, and 10% level. Standard errors are calculated by regressing the (standardized) preference parameter on each (standardized) individual characteristic, and are presented in parentheses. For the time discounting MPLs we estimate a monthly discount factor, δ. For the risk preference MPLs we use the certainty equivalent identified as the mid-point of the row in which a participant switched to the right hand side of the MPL. The survey contained two MPLs to measure each of WTA, WTP and time discounting. Correlations are estimated by stacking the two and clustering standard errors by participant.

Table E.4: Results of multivariate regressions in Table 2 are similar using categorical control variables.

	Risk Aversion $(1-\rho)$		Loss Aversion (λ)		Patience		Choice Consistency	
Cognitive Ability:								
Middle Tercile	-0.18**	-0.18**	-0.02	-0.03	0.19**	0.17**	0.26***	0.23***
	(0.078)	(0.078)	(0.082)	(0.080)	(0.081)	(0.080)	(0.085)	(0.083)
Top Tercile	-0.49***	-0.43***	0.36***	0.28***	0.48***	0.38***	0.40***	0.36***
	(0.071)	(0.075)	(0.083)	(0.084)	(0.080)	(0.082)	(0.078)	(0.075)
Age:								
36–50	0.04	0.04	-0.23**	-0.24**	0.15	0.13	0.09	0.02
	(0.096)	(0.096)	(0.106)	(0.104)	(0.110)	(0.108)	(0.097)	(0.095)
51–64	0.06	0.08	-0.32***	-0.35***	0.34***	0.30***	0.21**	0.14
	(0.092)	(0.093)	(0.092)	(0.092)	(0.098)	(0.097)	(0.094)	(0.094)
65+	0.01	-0.01	-0.26***	-0.28***	0.46***	0.48***	0.13	0.06
	(0.094)	(0.095)	(0.095)	(0.096)	(0.098)	(0.101)	(0.100)	(0.100)
Male	-0.12*	-0.11*	0.08	0.08	-0.11*	-0.10	-0.06	-0.05
	(0.062)	(0.061)	(0.067)	(0.065)	(0.066)	(0.063)	(0.067)	(0.064)
Education:								
Some College		-0.09		-0.01		0.30***		0.04
		(0.075)		(0.078)		(0.086)		(0.075)
4-year College		-0.09		0.16*		0.26***		0.13*
		(0.078)		(0.082)		(0.084)		(0.073)
Income:								
2nd Quartile		0.08		-0.03		-0.14		0.06
		(0.096)		(0.092)		(0.103)		(0.094)
3rd Quartile		0.01		0.19*		-0.12		-0.05
		(0.095)		(0.099)		(0.115)		(0.087)
4th Quartile		-0.25**		0.22**		0.10		-0.11
		(0.101)		(0.108)		(0.099)		(0.094)
Unreported		-0.06		0.34***		-0.01		-0.30**
		(0.115)		(0.120)		(0.108)		(0.120)
Married		0.03		0.07		0.07		0.18***
		(0.070)		(0.075)		(0.078)		(0.067)
N	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000

Notes: ***, **, * denote statistical significance at the 1%, 5%, and 10% level. All dependent variables are standardized. Robust standard errors are displayed in parentheses.

The results are similar when including all characteristics as categorical variables, to allow for potential non-monotonic relationships, as shown in Table E.4. Again, we see a strong relationship between cognitive ability and each of the DOSE-elicited parameters in each specification.

The results in Table 2 suggest that much of the correlation between education and both risk and loss aversion is explained by cognitive ability. To test that it is cognitive ability, and not one of the other controls, that weakens the association we carry out additional specifications adding the variables one at a time—see Table E.5. For each preference parameter, we start by adding education and income separately, then both together and, finally, add cognitive ability. It is only when cognitive ability is added that the magnitude of the coefficient with education diminishes significantly—suggesting that cognitive ability jointly determines educational outcomes and these two preferences.

Table E.5: Much of the relationship between education and risk preferences is explained by differences in cognitive ability.

	Risk Aversion $(1-\rho)$				Loss Aversion (λ)				Patience				Choice Consistency			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Cognitive Ability:																
Middle Tercile				-0.18**				-0.03				0.17**				0.23***
				(0.078)				(0.080)				(0.080)				(0.083)
Top Tercile				-0.43***				0.28***				0.38***				0.36***
				(0.075)				(0.084)				(0.082)				(0.075)
Education:																
Some College	-0.17**		-0.16**	-0.09	0.04		0.04	-0.01	0.35***		0.35***	0.30***	0.11		0.09	0.04
	(0.077)		(0.076)	(0.075)	(0.078)		(0.078)	(0.078)	(0.082)		(0.084)	(0.086)	(0.080)		(0.077)	(0.075)
4-year College	-0.28***		-0.20***	-0.09	0.32***		0.24***	0.16*	0.40***		0.36***	0.26***	0.23***		0.22***	0.13*
	(0.077)		(0.076)	(0.078)	(0.077)		(0.081)	(0.082)	(0.080)		(0.081)	(0.084)	(0.074)		(0.074)	(0.073)
Income:																
2nd Quartile		0.04	0.06	0.08		0.00	-0.02	-0.03		-0.07	-0.12	-0.14		0.10	0.08	0.06
		(0.099)	(0.098)	(0.096)		(0.091)	(0.091)	(0.092)		(0.108)	(0.105)	(0.103)		(0.094)	(0.094)	(0.094)
3rd Quartile		-0.07	-0.01	0.01		0.27***	0.21**	0.19*		0.00	-0.10	-0.12		0.02	-0.03	-0.05
		(0.097)	(0.096)	(0.095)		(0.099)	(0.098)	(0.099)		(0.118)	(0.116)	(0.115)		(0.089)	(0.089)	(0.087)
Top Quartile		-0.37***	-0.30***	-0.25**		0.36***	0.26**	0.22**		0.27***	0.14	0.10		0.02	-0.07	-0.11
		(0.100)	(0.101)	(0.101)		(0.102)	(0.107)	(0.108)		(0.101)	(0.104)	(0.099)		(0.089)	(0.094)	(0.094)
Unreported		-0.08	-0.07	-0.06		0.38***	0.36***	0.34***		0.01	-0.00	-0.01		-0.28**	-0.29**	-0.30**
		(0.115)	(0.115)	(0.115)		(0.122)	(0.122)	(0.120)		(0.111)	(0.110)	(0.108)		(0.125)	(0.125)	(0.120)
Married		0.03	0.02	0.03		0.06	0.07	0.07		0.06	0.07	0.07		0.18***	0.19***	0.18***
		(0.071)	(0.071)	(0.070)		(0.075)	(0.074)	(0.075)		(0.081)	(0.080)	(0.078)		(0.068)	(0.069)	(0.067)
N	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000

Notes: ***, **, * denote statistical significance at the 1%, 5%, and 10% level. All dependent variables are standardized. Robust standard errors are displayed in parentheses.

Tables E.6–E.8 show that we identify similar correlations when restricting the sample to participants most likely to be paying attention—those classified as having high choice consistency, or with above median response time in either the DOSE modules or the survey as a whole.

Table E.6: Results in Table 2 are similar among participants with high choice consistency.

	Risk Aversion $(1-\rho)$		Loss Aversion (λ)		Patience (δ)		Choice Consistency (μ)	
	Univariate Correlation	Multivariate Regression	Univariate Correlation	Multivariate Regression	Univariate Correlation	Multivariate Regression	Univariate Correlation	Multivariate Regression
Cognitive Ability	-0.18***	-0.13***	0.21***	0.16***	0.16***	0.17***	-0.04	-0.01
	(0.039)	(0.043)	(0.041)	(0.044)	(0.044)	(0.039)	(0.038)	(0.040)
Education	-0.12***	-0.06	0.14***	0.05	0.16***	0.11*	-0.11***	-0.12***
	(0.048)	(0.049)	(0.043)	(0.044)	(0.057)	(0.058)	(0.045)	(0.050)
Income	-0.09*	-0.04	0.15***	0.10*	0.10**	-0.00	0.01	0.03
	(0.052)	(0.059)	(0.040)	(0.050)	(0.045)	(0.050)	(0.040)	(0.045)
Male	-0.12***	-0.09*	0.05	-0.00	-0.03	-0.05	-0.03	-0.03
	(0.046)	(0.046)	(0.044)	(0.043)	(0.052)	(0.047)	(0.044)	(0.043)
Age	0.03	0.01	-0.08*	-0.08	0.21***	0.22***	-0.04	-0.06
	(0.047)	(0.049)	(0.047)	(0.050)	(0.057)	(0.056)	(0.046)	(0.046)
Married	0.01	0.02	0.03	0.02	0.08	0.03	0.06	0.06
	(0.047)	(0.053)	(0.044)	(0.049)	(0.054)	(0.057)	(0.045)	(0.046)
N	1,033	1,033	1,033	1,033	1,033	1,033	1,033	1,033

Notes: ***, **, * denote statistical significance at the 1%, 5%, and 10% level. The table repeats the specifications in Table 2 restricting the sample to those with above median choice consistency. Robust standard errors, in parentheses, come from a standardized regression. The first, third, fifth, and seventh columns report univariate correlations. The second, fourth, sixth, and eighth columns report the coefficient from a multivariate regression.

Table E.7: Results in Table 2 are similar among participants with above median response time in DOSE modules.

	Risk Aversion $(1-\rho)$		Loss Aversion (λ)		Patience (δ)		Choice Consistency (μ)	
	Univariate Correlation	Multivariate Regression	Univariate Correlation	Multivariate Regression	Univariate Correlation	Multivariate Regression	Univariate Correlation	Multivariate Regression
Cognitive Ability	-0.21***	-0.17***	0.26***	0.23***	0.18***	0.20***	0.16***	0.15***
	(0.039)	(0.044)	(0.037)	(0.039)	(0.041)	(0.040)	(0.036)	(0.039)
Education	-0.14***	-0.08**	0.13***	0.04	0.16***	0.09**	0.09*	0.03
	(0.039)	(0.043)	(0.041)	(0.042)	(0.045)	(0.045)	(0.047)	(0.047)
Income	-0.07	0.00	0.10***	0.00	0.10*	-0.01	0.08*	-0.00
	(0.053)	(0.061)	(0.044)	(0.051)	(0.051)	(0.054)	(0.045)	(0.053)
Male	-0.10***	-0.06	0.08*	0.02	-0.04	-0.08*	-0.01	-0.04
	(0.044)	(0.042)	(0.042)	(0.042)	(0.050)	(0.045)	(0.046)	(0.048)
Age	0.07	0.04	-0.04	-0.02	0.19***	0.21***	0.04	0.04
	(0.045)	(0.046)	(0.047)	(0.045)	(0.056)	(0.053)	(0.057)	(0.057)
Married	-0.03	-0.00	0.11***	0.08*	0.11**	0.05	0.12***	0.10*
	(0.044)	(0.051)	(0.042)	(0.045)	(0.050)	(0.051)	(0.047)	(0.051)
N	1,012	1,012	1,012	1,012	1,012	1,012	1,012	1,012

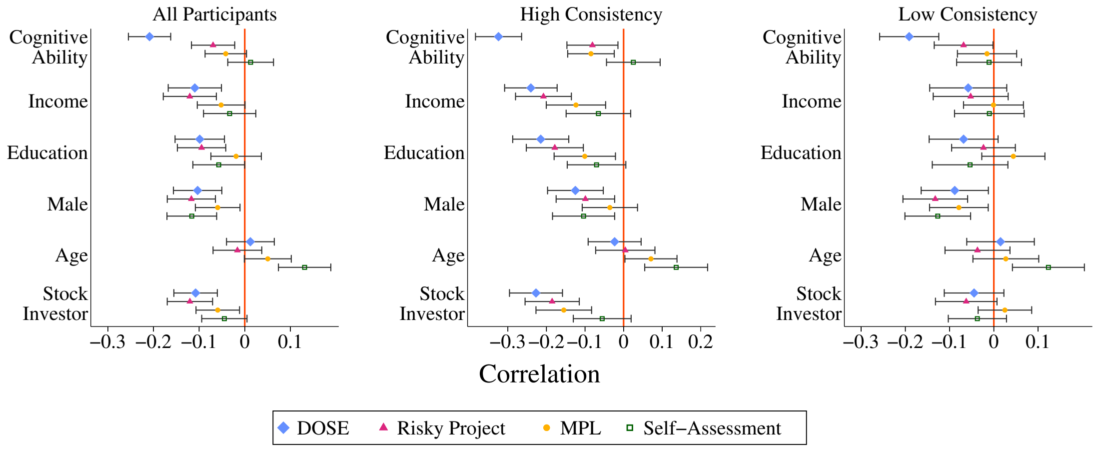
Notes: ***, **, * denote statistical significance at the 1%, 5%, and 10% level. The table repeats the specifications in Table 2 restricting the sample to those with above median response time on the DOSE modules. Robust standard errors, in parentheses, come from a standardized regression. The first, third, fifth, and seventh columns report univariate correlations. The second, fourth, sixth, and eighth columns report the coefficient from a multivariate regression.

Table E.8: Results in Table 2 are similar among participants with above median response time in the survey as a whole.

	Risk Aversion $(1-\rho)$		Loss Aversion (λ)		Patience (δ)		Choice Consistency (μ)	
	Univariate Correlation	Multivariate Regression	Univariate Correlation	Multivariate Regression	Univariate Correlation	Multivariate Regression	Univariate Correlation	Multivariate Regression
Cognitive Ability	-0.21***	-0.16***	0.22***	0.20***	0.17***	0.18***	0.18***	0.16***
	(0.039)	(0.044)	(0.039)	(0.043)	(0.042)	(0.042)	(0.035)	(0.038)
Education	-0.14***	-0.07	0.07	-0.03	0.15***	0.08*	0.10***	0.03
	(0.044)	(0.047)	(0.042)	(0.044)	(0.043)	(0.045)	(0.039)	(0.043)
Income	-0.11**	-0.04	0.12***	0.05	0.08*	-0.02	0.12***	0.03
	(0.051)	(0.060)	(0.043)	(0.052)	(0.050)	(0.055)	(0.040)	(0.043)
Male	-0.12***	-0.08*	0.08*	0.03	-0.04	-0.07	0.02	-0.02
	(0.045)	(0.043)	(0.043)	(0.043)	(0.051)	(0.048)	(0.043)	(0.043)
Age	0.03	0.02	-0.08	-0.06	0.18***	0.19***	-0.00	0.00
	(0.047)	(0.047)	(0.047)	(0.046)	(0.057)	(0.057)	(0.046)	(0.047)
Married	-0.02	0.03	0.09**	0.06	0.07	0.03	0.11***	0.08*
	(0.044)	(0.052)	(0.043)	(0.047)	(0.051)	(0.053)	(0.044)	(0.046)
N	993	993	993	993	993	993	993	993

Notes: ***, **, * denote statistical significance at the 1%, 5%, and 10% level. The table repeats the specifications in Table 2 restricting the sample to those with above median response time on the survey as a whole. Robust standard errors, in parentheses, come from a standardized regression. The first, third, fifth, and seventh columns report univariate correlations. The second, fourth, sixth, and eighth columns report the coefficient from a multivariate regression.

Figure E.7: The patterns of correlations in Figure 8 are similar using an alternative definition of choice consistency.



Notes: Figure repeats the analysis in Figure 8, using the choice consistency variable identified from the full 20-question DOSE sequence. See notes to that figure for more details.

E.5 Choice Consistency and Response Time

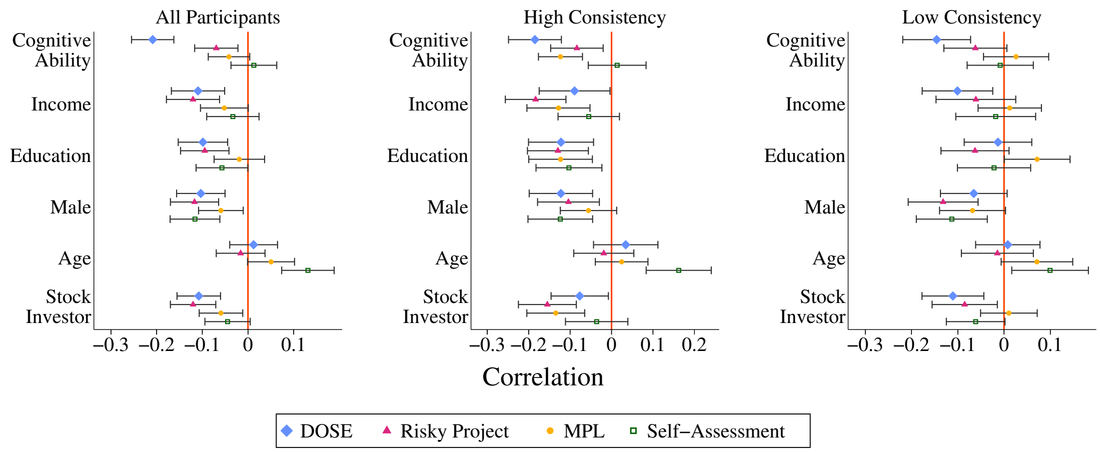
Figure E.7 shows that the findings in Figure 8 are robust to using the choice consistency variable identified from both DOSE modules, rather than simply the 10-question time preference module. The left-hand panel is identical to Figure 8, but the middle- and right-hand panel now split the same according to this alternative definition of consistent choice. The patterns are similar, but quite not as clear, consistent with the time preference module providing more informative estimates of μ.

Figure E.8 shows that the findings in Figure 8 are similar when using non-parameteric measures of risk aversion from the MPL and risky project elicitations.

We now show that that controlling for choice consistency helps identify a pattern of correlations even when restricting the sample to those answering very fast—and so who might be thought to be paying little attention. In the left hand panel of Figure E.9 we show the pattern of correlations restricting the sample to first those answering the risk MPL module quickly and in the right hand panel we present the correlations for those answering the whole survey quickly (quickly being defined as below the respective median). In both cases we compare the correlations for all participants to those in the high consistency group.

In both panels there is more evidence of correlations after restricting the sample to high consistency participants. The magnitude of the correlations in frequently higher, and several emerge as statistically significant once only high consistency participants are considered. The magnitude of the correlations is, in fact, similar to those in Figure 8, although the standard errors are larger (explained by the fact the sample is half as large). The choice consistency measure appears, then, to be distinguishing participants that answer accurately but rapidly—

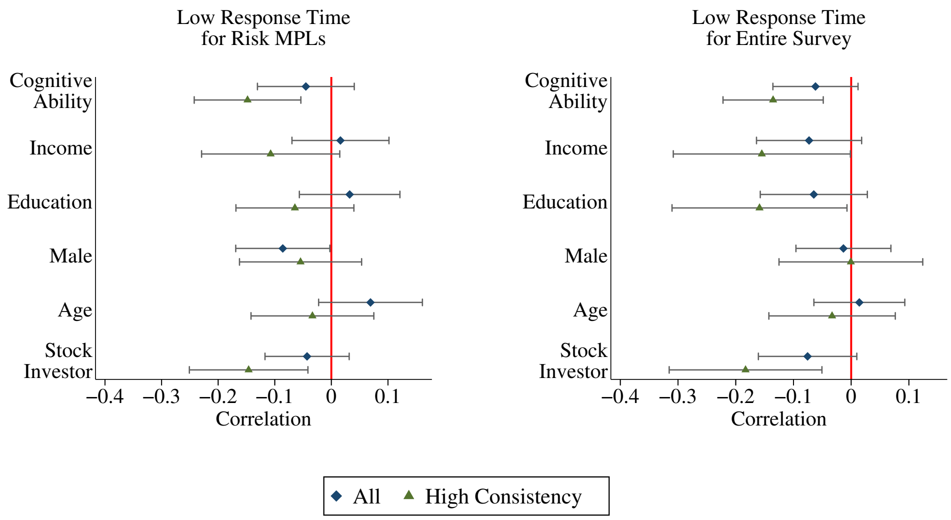
Figure E.8: The patterns of correlations in Figure 8 are similar using non-parametric measures for MPL and risky project elicitations.



Notes: Figure repeats the analysis in Figure 8, using non-parametric measures of risk aversion from the MPL and risky project elicitations. See notes to that figure for more details.

whose responses include meaningful information—from those that answer quickly due to a lack of care or attention.

Figure E.9: Accounting for choice consistency leads to a clearer pattern of correlations even among participants with low response times.



Notes: The left panel includes only participants below the median response time on the risk MPL module. The right panel includes only participants below the median response time on the entire survey. “High Consistency” refers to those with choice consistency above the median. The survey contained two MPL measures of risk preference. Correlations are estimated by stacking the two and clustering standard errors by participant.

F Screenshots

This subsection contains screenshots of all the types of questions analyzed in this paper. Full design documents and screenshots can be found at hss.caltech.edu/∼snowberg/wep.html.

Figure F.10: DOSE Risk/Loss Aversion Instruction Screen

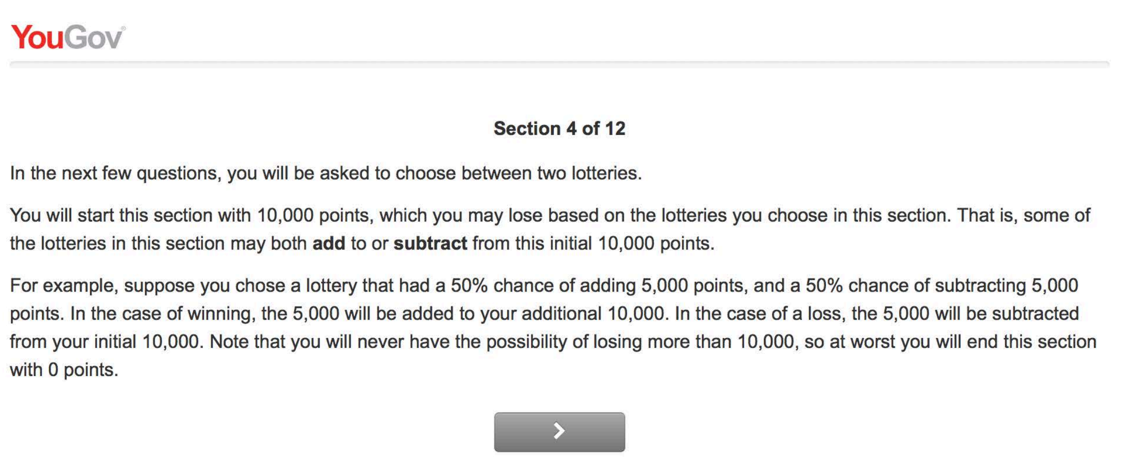


Figure F.11: DOSE Risk/Loss Aversion Example Question: Gains Only

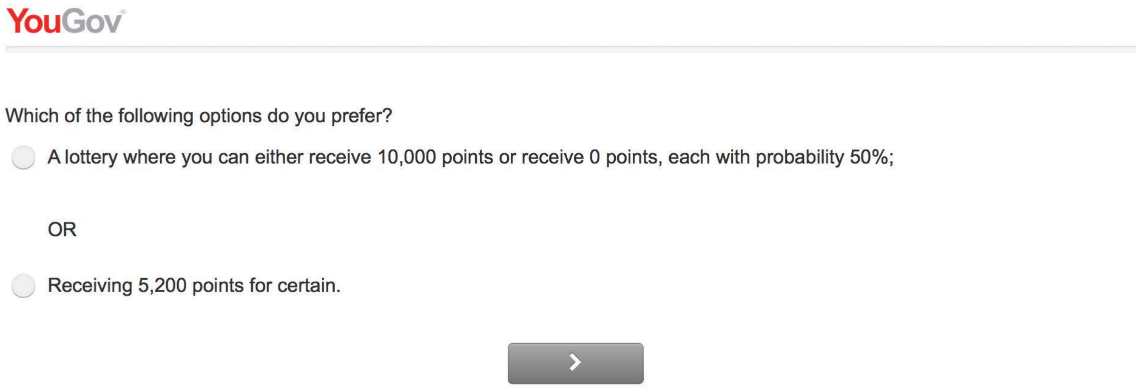


Figure F.12: DOSE Risk/Loss Aversion Example Question: Both Gains and Losses

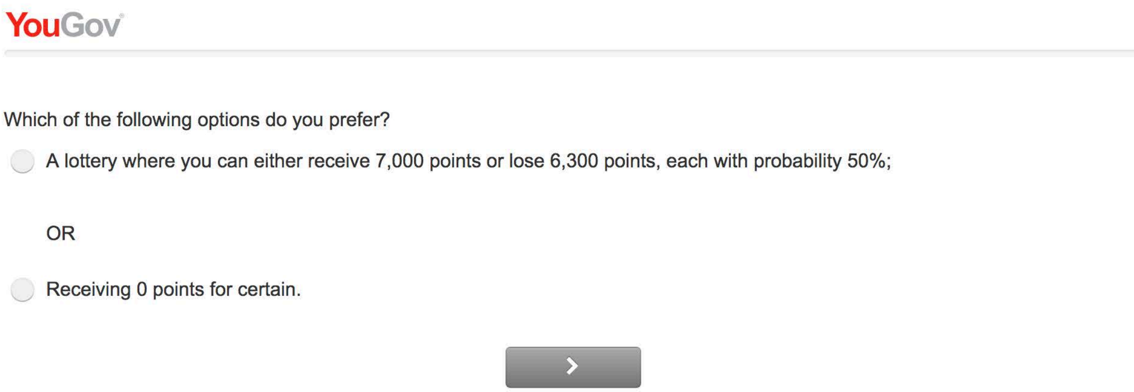


Figure F.13: DOSE Discounting Instruction Screen

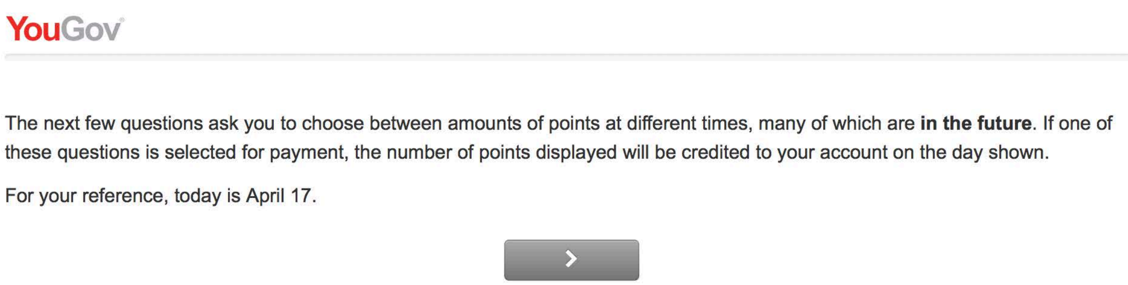
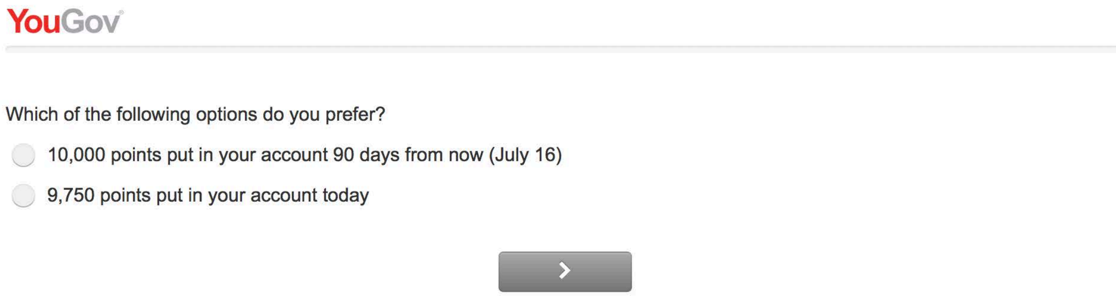


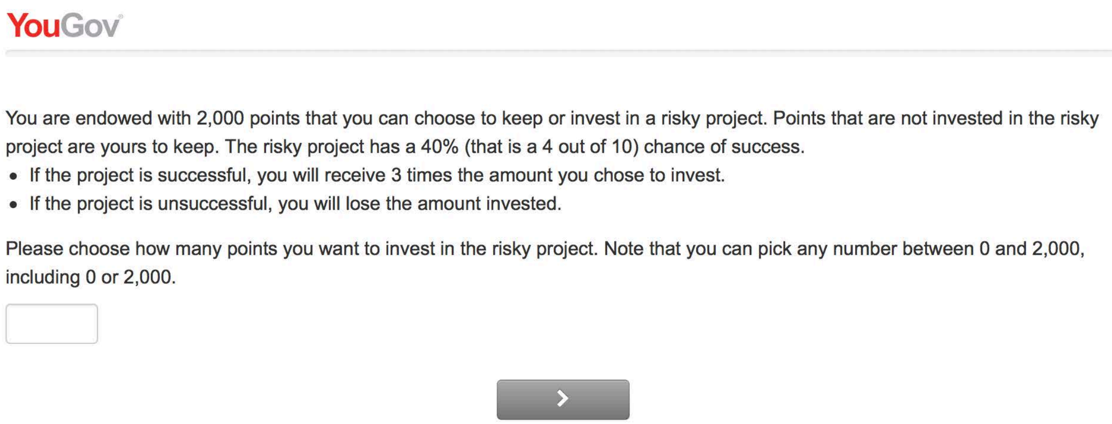
Figure F.14: DOSE Discounting Example Question


YouGov

Which of the following options do you prefer?

- 10,000 points put in your account 90 days from now (July 16)
- 9,750 points put in your account today

Figure F.15: Risky Project Question


YouGov

You are endowed with 2,000 points that you can choose to keep or invest in a risky project. Points that are not invested in the risky project are yours to keep. The risky project has a 40% (that is a 4 out of 10) chance of success.

- If the project is successful, you will receive 3 times the amount you chose to invest.
- If the project is unsuccessful, you will lose the amount invested.

Please choose how many points you want to invest in the risky project. Note that you can pick any number between 0 and 2,000, including 0 or 2,000.

Figure F.16: Multiple Price List to Measure Risk Aversion

YouGov

For this question, you **have been given 8,000 points**. You will be offered the opportunity to exchange some of these points for a lottery ticket. This lottery ticket has a **50% chance** of paying you **8,000 points**, and a **50% chance** of paying **2,000 points**.

For example, if you choose to pay 3,000 points for a lottery ticket, and this question is chosen for payment, you will:

- Pay 3,000 points for the lottery ticket
- Keep 5,000 points for yourself
- Earn whatever proceeds you get from the lottery ticket (if any)

For each row in the table below, which option would you prefer?

☑ Keep 8,000 points	or	☐ Buy the lottery ticket for 8,000 points and keep the remaining 0 points
☐ Keep 8,000 points	or	☐ Buy the lottery ticket for 6,000 points and keep the remaining 2,000 points
☐ Keep 8,000 points	or	☐ Buy the lottery ticket for 5,500 points and keep the remaining 2,500 points
☐ Keep 8,000 points	or	☐ Buy the lottery ticket for 5,250 points and keep the remaining 2,750 points
☐ Keep 8,000 points	or	☐ Buy the lottery ticket for 5,000 points and keep the remaining 3,000 points
☐ Keep 8,000 points	or	☐ Buy the lottery ticket for 4,750 points and keep the remaining 3,250 points
☐ Keep 8,000 points	or	☐ Buy the lottery ticket for 4,500 points and keep the remaining 3,500 points
☐ Keep 8,000 points	or	☐ Buy the lottery ticket for 4,250 points and keep the remaining 3,750 points
☐ Keep 8,000 points	or	☐ Buy the lottery ticket for 4,000 points and keep the remaining 4,000 points
☐ Keep 8,000 points	or	☐ Buy the lottery ticket for 3,750 points and keep the remaining 4,250 points
☐ Keep 8,000 points	or	☐ Buy the lottery ticket for 3,500 points and keep the remaining 4,500 points
☐ Keep 8,000 points	or	☐ Buy the lottery ticket for 3,250 points and keep the remaining 4,750 points
☐ Keep 8,000 points	or	☐ Buy the lottery ticket for 3,000 points and keep the remaining 5,000 points
☐ Keep 8,000 points	or	☐ Buy the lottery ticket for 2,500 points and keep the remaining 5,500 points
☐ Keep 8,000 points	or	☑ Buy the lottery ticket for 2,000 points and keep the remaining 6,000 points

Reset Autofill

Review the instructions

Figure F.17: Multiple Price List to Estimate Discount Rate

YouGov

For each row in the table below, which option would you prefer?

☑ 6,000 points in 45 days (June 1)	or	☐ 0 points today
☐ 6,000 points in 45 days (June 1)	or	☐ 1,000 points today
☐ 6,000 points in 45 days (June 1)	or	☐ 2,000 points today
☐ 6,000 points in 45 days (June 1)	or	☐ 3,000 points today
☐ 6,000 points in 45 days (June 1)	or	☐ 3,500 points today
☐ 6,000 points in 45 days (June 1)	or	☐ 4,000 points today
☐ 6,000 points in 45 days (June 1)	or	☐ 4,500 points today
☐ 6,000 points in 45 days (June 1)	or	☐ 5,000 points today
☐ 6,000 points in 45 days (June 1)	or	☐ 5,500 points today
☐ 6,000 points in 45 days (June 1)	or	☐ 5,600 points today
☐ 6,000 points in 45 days (June 1)	or	☐ 5,700 points today
☐ 6,000 points in 45 days (June 1)	or	☐ 5,800 points today
☐ 6,000 points in 45 days (June 1)	or	☐ 5,900 points today
☐ 6,000 points in 45 days (June 1)	or	☐ 5,950 points today
☐ 6,000 points in 45 days (June 1)	or	☐ 5,975 points today
☐ 6,000 points in 45 days (June 1)	or	☐ 6,000 points today
☐ 6,000 points in 45 days (June 1)	or	☐ 6,100 points today

Reset | Autofill

Review the instructions

Figure F.18: Multiple Price List Instruction Screen

YouGov

This survey often uses a special type of question. We want to help you answer these questions **quickly** and **accurately**.

This special type of question has many similar choices, as in the example below. The options on the left are always the same, while those on the right change — getting better and better.

If a question like this is picked for payment, **one row** will be selected, and you will be paid according to the choice **you made in that row**. It is important that your answers in each row **are accurate** so you will get the payment **you want**.

You will see a screen that looks like this.

☑ 5,000 points	or	☐ 0 points
☐ 5,000 points	or	☐ 1,000 points
☐ 5,000 points	or	☐ 2,000 points
☐ 5,000 points	or	☐ 3,000 points
☐ 5,000 points	or	☐ 4,500 points
☐ 5,000 points	or	☐ 5,500 points
☐ 5,000 points	or	☐ 6,000 points
☐ 5,000 points	or	☐ 7,000 points
☐ 5,000 points	or	☐ 8,000 points
☐ 5,000 points	or	☐ 9,000 points
☐ 5,000 points	or	☑ 10,000 points

Figure F.19: Multiple Price List Practice Screen

YouGov

To answer these types of questions **quickly** and **accurately** we suggest you:

1. Start by looking at the **top row**, and think carefully about each row in turn.
2. For each row where you **prefer the option on the left** over the option on the right, check the box on the left hand side.
3. When you find the **first question where you prefer the option on the right** over the option on the left, check the box on the right.
4. Notice that the option on the right is always better as you go down the list. This means that after you choose one option on the right, you should choose the option on the right for all rows below. Your answers should therefore "cross over" from left to right **only once**.
5. Once you have filled in the "cross over" point you may hit the Autofill button to fill in the rest of the chart faster. Alternatively, you may check every box manually.

All rows must have a box checked for you to continue to the next page

If you need to start over at any point, hit the **Reset** button to clear out all of the checkmarks.

Example question: For each row in the table below, which option would you prefer?

☑ 5,000 points	or	☐ 0 points
☐ 5,000 points	or	☐ 1,000 points
☐ 5,000 points	or	☐ 2,000 points
☐ 5,000 points	or	☐ 3,000 points
☐ 5,000 points	or	☐ 4,500 points
☐ 5,000 points	or	☐ 5,500 points
☐ 5,000 points	or	☐ 6,000 points
☐ 5,000 points	or	☐ 7,000 points
☐ 5,000 points	or	☐ 8,000 points
☐ 5,000 points	or	☐ 9,000 points
☐ 5,000 points	or	☑ 10,000 points

Reset Autofill

Review the instructions

Figure F.20: Multiple Price List Practice Screen with Error Message

YouGov

To answer these types of questions **quickly** and **accurately** we suggest you:

1. Start by looking at the **top row**, and think carefully about each row in turn.
2. For each row where you **prefer the option on the left** over the option on the right, check the box on the left hand side.
3. When you find the **first question where you prefer the option on the right** over the option on the left, check the box on the right.
4. Notice that the option on the right is always better as you go down the list. This means that after you choose one option on the right, you should choose the option on the right for all rows below. Your answers should therefore "cross over" from left to right **only once**.
5. Once you have filled in the "cross over" point you may hit the Autofill button to fill in the rest of the chart faster. Alternatively, you may check every box manually.

All rows must have a box checked for you to continue to the next page

If you need to start over at any point, hit the **Reset** button to clear out all of the checkmarks.

Example question: For each row in the table below, which option would you prefer?

You have indicated in row 2 that you prefer 1,000 points to 5,000 points. But 1,000 points is less than 5,000 points, which means you would get more by selecting 5,000 points. Please correct this.

In all the other questions on this survey, there is no right or wrong answer. However, you should make sure that you select the option that you prefer on **each line**.

☑ 5,000 points	or	☐ 0 points
☐ 5,000 points	or	☑ 1,000 points
☐ 5,000 points	or	☑ 2,000 points
☐ 5,000 points	or	☑ 3,000 points
☐ 5,000 points	or	☑ 4,500 points
☐ 5,000 points	or	☑ 5,500 points
☐ 5,000 points	or	☑ 6,000 points
☐ 5,000 points	or	☑ 7,000 points
☐ 5,000 points	or	☑ 8,000 points
☐ 5,000 points	or	☑ 9,000 points
☐ 5,000 points	or	☑ 10,000 points

Reset

Autofill

Review the instructions

Figure F.21: Qualitative Self-Assessment: Risk Preferences

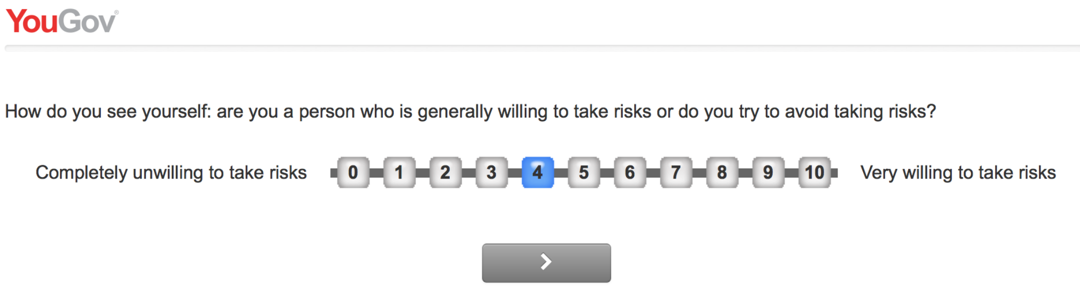
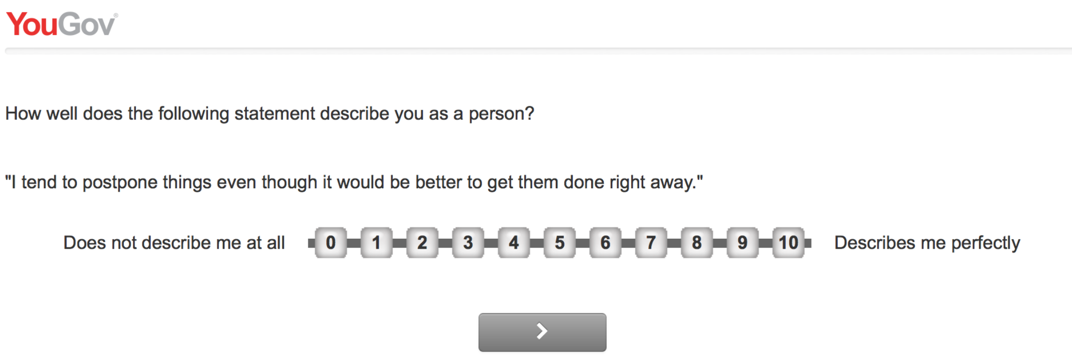


Figure F.22: Qualitative Self-Assessment: Time Preferences



References

Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutström, "Eliciting Risk and Time Preferences," *Econometrica*, 2008, *76* (3), 583–618.

Andreoni, James and Charles Sprenger, "Estimating Time Preferences from Convex Budgets," *The American Economic Review*, 2012, *102* (7), 3333–3356.

Apesteguia, Jose and Miguel A. Ballester, "Monotone Stochastic Choice Models: The Case of Risk and Time Preferences," *Journal of Political Economy*, 2018, *126* (1), 74–106.

Bagaïni, Alexandra, Yunrui Liu, Arzie Bajrami, Gayoung Son, Loreen Tisdall, and Rui Mata, "Aging and Economic Preferences: Cumulative meta-analyses of age differences in risk, time, social, and effort preferences," *The Journals of Gerontology: Series B*, 2023, *78* (7), 1122–1135.

Barsky, Robert B. F. Thomas Juster, Miles S. Kimball, and Matthew D. Shapiro, "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study," *The Quarterly Journal of Economics*, 1997, *112* (2), 537–579.

Benjamin, Daniel J., Sebastian A. Brown, and Jesse M. Shapiro, "Who Is 'Behavioral'? Cognitive Ability and Anomalous Preferences," *Journal of the European Economic Association*, 2013, *11* (6), 1231–1255.

Burks, Stephen V., Jeffrey P. Carpenter, Lorenz Goette, and Aldo Rustichini, "Cognitive Skills Affect Economic Preferences, Strategic Behavior, and Job Attachment," *Proceedings of the National Academy of Sciences*, 2009, *106* (19), 7745–7750.

Chapman, Jonathan, Erik Snowberg, Stephanie Wang, and Colin F. Camerer, "Looming Large or Seeming Small? Attitudes Towards Losses in a Representative Sample," *Review of Economic Studies*, forthcoming.

_ , Mark Dean, Pietro Ortoleva, Erik Snowberg, and Colin Camerer, "Willingness to Accept, Willingness to Pay, and Loss Aversion," 2023. NBER Working Paper #30836.

Chuang, Yating and Laura Schechter, "Stability of Experimental and Survey Measures of Risk, Time, and Social Preferences: A Review and Some New Results," *Journal of Development Economics*, 2015, *117*, 151–170.

Clay, Summer N. Allison M. Harris, and Catherine L. Reed, "Loss Aversion Affects Inhibitory Processes for Reward and Loss as Indicated by Inhibition of Return," *Journal of Vision*, 2016, *16* (12), 95.

_ , John A. Clithero, Alison M. Harris, and Catherine L. Reed, "Loss Aversion Reflects Information Accumulation, Not Bias: A Drift-Diffusion Model Study," *Frontiers in Psychology*, 2017, *8*, 1708.

Condon, David M. and William Revelle, "The International Cognitive Ability Resource: Development and Initial Validation of a Public-Domain Measure," *Intelligence*, 2014, *43*, 52–64.

Dave, Chetan, Catherine C. Eckel, Cathleen A. Johnson, and Christian Rojas, "Eliciting Risk Preferences: When is Simple Better?," *Journal of Risk and Uncertainty*, 2010, *41* (3), 219–243.

Dean, Mark and Anja Sautmann, "Credit Constraints and the Measurement of Time Preferences," 2014. Columbia University, *mimeo.*

DellaVigna, Stefano and M. Daniele Paserman, "Job Search and Impatience," *Journal* *of Labor Economics*, 2005, *23* (3), 527–588.

Dohmen, Thomas, Armin Falk, Armin Huffman, and Uwe Sunde, "Are Risk Aversion and Impatience Related to Cognitive Ability?," *The American Economic Review*, 2010, *100* (3), 1238–1260.

Eckel, Catherine C. and Philip J. Grossman, "Sex Differences and Statistical Stereotyping in Attitudes toward Financial Risk," *Evolution and Human Behavior*, 2002, *23* (4), 281–295.

_ and _ , "Men, Women and Risk Aversion: Experimental Evidence," in Charles R. Plott and Vernon L. Smith, eds., *Handbook of Experimental Economics Results*, Vol. 1, North-Holland, 2008, pp. 1061–1073.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde, "Global Evidence on Economic Preferences," *The Quarterly Journal of Economics*, 2018, pp. 1645—-1692.

_ , _ , _ , David Huffman, and Uwe Sunde, "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences," *Management Science*, 2023, *69* (4), 1935–1950.

Frederick, Shane, "Cognitive Reflection and Decision Making," *Journal of Economic Perspectives*, 2005, *19* (4), 25–42.

Frydman, Cary, Colin Camerer, Peter Bossaerts, and Antonio Rangel, "MAOA-L Carriers are Better at Making Optimal Financial Decisions under Risk," *Proceedings of the Royal Society of London B: Biological Sciences*, 2011, *278* (1714), 2053–2059.

Gneezy, Uri and Jan Potters, "An Experiment on Risk Taking and Evaluation Periods," *The Quarterly Journal of Economics*, 1997, *112* (2), 631–645.

Harrison, Glenn W. Morten I. Lau, and E. Elisabet Rutström, "Estimating Risk Attitudes in Denmark: A Field Experiment," *The Scandinavian Journal of Economics*, 2007, *109* (2), 341–368.

Holt, Charles A. and Susan K. Laury, "Risk Aversion and Incentive Effects," *The American Economic Review*, 2002, *92* (5), 1644–1655.

Imai, Taisuke and Colin F. Camerer, "Estimating Time Preferences from Budget Set Choices Using Optimal Adaptive Design," 2018. Ludwig Maximilian University of Munich, *mimeo.*

Johnson, Cathleen, Aurélien Baillon, Han Bleichrodt, Zhihua Li, Dennie Van Dolder, and Peter P Wakker, "PRINCE: An Improved Method for Measuring Incentivized Preferences," *Journal of Risk and Uncertainty*, 2021, *62* (1), 1–28.

Kirby, Kris N. Ricardo Godoy, Victoria Reyes-Garcıa, Elizabeth Byron, Lilian Apaza, William Leonard, Eddy Perez, Vincent Vadez, and David Wilkie, "Correlates of Delay-Discount Rates: Evidence from Tsimane'Amerindians of the Bolivian Rain Forest," *Journal of Economic Psychology*, 2002, *23* (3), 291–316.

Krajbich, Ian, Colin Camerer, and Antonio Rangel, "Exploring the Scope of Neurometrically Informed Mechanism Design," *Games and Economic Behavior*, 2017, *101*, 49–62.

Lilleholt, Lau, "Cognitive ability and risk aversion: A systematic review and meta analysis," *Judgment and Decision making*, 2019, *14* (3), 234–279.

Mata, Rui, Renato Frey, David Richter, Jürgen Schupp, and Ralph Hertwig, "Risk Preference: A View from Psychology," *Journal of Economic Perspectives*, 2018, *32*

(2), 155–72.

Meier, Stephan and Charles Sprenger, "Present-biased preferences and credit card borrowing," *American Economic Journal: Applied Economics*, 2010, *2* (1), 193–210.

_ and _ , "Temporal Stability of Time Preferences," *Review of Economics and Statistics*, 2015, *97* (2), 273–286.

Olschewski, Sebastian, Jörg Rieskamp, and Benjamin Scheibehenne, "Taxing cognitive capacities reduces choice consistency rather than preference: A model-based test.," *Journal of Experimental Psychology: General*, 2018, *147* (4), 462.

Ray, Debajyoti, Daniel Golovin, Andreas Krause, and Colin Camerer, "Bayesian Rapid Optimal Adaptive Design (BROAD): Method and Application Distinguishing Models of Risky Choice," 2012. Caltech, *mimeo.*

Shamosh, Noah A and Jeremy R Gray, "Delay Discounting and Intelligence: A Meta-Analysis," *Intelligence*, 2008, *36* (4), 289–305.

Sokol-Hessner, Peter, Ming Hsu, Nina G. Curley, Mauricio R. Delgado, Colin F. Camerer, and Elizabeth A. Phelps, "Thinking Like a Trader Selectively Reduces Individuals' Loss Aversion," *Proceedings of the National Academy of Sciences*, 2009, *106* (13), 5035–5040.

Tanaka, Tomomi, Colin F Camerer, and Quang Nguyen, "Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam," *The American Economic Review*, 2010, *100* (1), 557–71.

Urminsky, Oleg and Gal Zauberman, "The Psychology of Intertemporal Preferences," in Gideon Keren and George Wu (Editor), eds., *Handbook of Judgment and Decision Making*, Vol. 2, Wiley-Blackwell, 2016, pp. 141–181.

Wang, Stephanie W. Michelle Filiba, and Colin Camerer, "Dynamically Optimized Sequential Experimentation (DOSE) for Estimating Economic Preference Parameters," 2010. Caltech, *mimeo.*

Wölbert, Eva and Arno Riedl, "Measuring Time and Risk Preferences: Reliability, Stability, Domain Specificity," 2013. *CESifo Working Paper Series No. 4339.*