ECONtribute
MARKETS & PUBLIC POLICY

ECONtribute
Discussion Paper No. 350

Separating Preferences from Endogenous Effort and Cognitive Noise in Observed Decisions

Christian Belzil

Tomáš Jagelka

December 2024

www.econtribute.de

UNIVERSITÄT BONN

UNIVERSITÄT ZU KÖLN

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1-390838866 is gratefully acknowledged.

Separating Preferences from Endogenous Effort and Cognitive Noise in Observed Decisions*

Authors:

Christian BELZIL† and Tomáš JAGELKA‡

Abstract

We develop a micro-founded framework for accounting for individuals' effort and cognitive noise which confound estimates of preferences based on observed behavior. Using a large-scale experimental dataset we estimate that failure to properly account for decision errors due to (rational) inattention on a more complex, but commonly used, task design biases estimates of risk aversion by 50% for the median individual. Effort propensities recovered from preference elicitation tasks generalize to other settings and predict performance on an OECD-sponsored achievement test used to make international comparisons. Furthermore, accounting for endogenous effort allows us to empirically reconcile competing models of discrete choice.

December 16, 2024

*Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

We would like to thank Larbi Alaoui, Jose Apesteguia, Miguel Ballester, Mark Dean, James Heckman, Nicolas Jacquemet, Erzo F.P. Luttmer, Jawwad Noor, Antonio Penta, Philipp Schirmer, Nikolaus Schweizer, Douglas Staiger, Robert Sugden, John Rust, and Erik Wengström for helpful discussions, comments, and suggestions. We also thank participants at seminars at the University of Bonn, CERGE-EI, the Economic University of Bratislava, Laval University, MIT, Universitat Pompeu Fabra, Université Sorbonne at Paris 1 and Paris School of Economics, at the CIRANO workshop in Behavioral and Experimental Economics for Innovative Policy Making, the Czech Economic Society and Slovak Economic Association Meeting, the European Economic Association Congress, the ESA Junior Faculty Webinar, the ESA Global Meetings, the IAREP Early Career Workshop, the IAREP/SABE 2024 Joint Meeting, the North American Summer Meeting of the Econometric Society, and at the Preferences, Traits, and Skills workshop for valuable feedback. This work is supported by ERC grant: FELICITAS, 101165518 and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1 – 390838866.

†CREST, CNRS, Paris Polytechnic Institute, IZA, and CIRANO. Email: christian.belzil@polytechnique.edu.

‡University of Bonn, Dartmouth College, CREST-Ensae, and IZA. Email: tjagelka@uni-bonn.de.

1 Introduction

Preferences, like skills and other latent personal attributes, are key drivers of inequalities in life outcomes. Themselves unobserved, they need to be inferred from observed behavior. Heckman, Jagelka, and Kautz (2021) clarify that performance on *any* task is a function of multiple preferences, skills, and also of effort, which in turn depends on task-specific incentives. Careful experimental and survey design attempts to isolate the impact of a particular preference (or of another latent attribute of interest) on observed decisions. However, cognitive noise and decision mistakes due to inattention remain as potential sources of bias. Our main contribution is to develop and estimate a micro-founded stochastic choice model which separates the signal on preferences in observed choices from endogenous effort and cognitive noise.[1] It allows us to (i) de-bias estimates of risk preferences, (ii) uncover an individual-specific tendency to exert effort which generalizes beyond experimental settings, and (iii) reconcile competing models of discrete choice.

As demonstrated in the complexity literature (e.g., Gabaix and Graeber, 2023), the frequency of individual decision errors is linked to the inherent level of cognitive difficulty of an experimental task. As a consequence, the observed pattern of individual decisions is contaminated and may induce statistical bias when estimating structural preference parameters.[2] This can have large policy implications given that, for example, the Netherlands now legally require pension funds to measure the risk attitudes of their members in a quantitative way (see, e.g., Goossens et al., 2023). Our model implies a general relationship between elicitation task complexity and bias in preference estimates (risk, time, social, etc.) when the relationship between task complexity and individuals' propensity to exert effort is ignored. We show that choices on the more complex of two popular task designs used in the literature for eliciting risk preferences yield estimates of risk aversion biased by approximately 50% for the median individual when effort is not properly accounted for. We provide a simple general formula for predicting bias in preference estimates at the individual level and demonstrate its effectiveness even in preference elicitation tasks with many choice options.

By incorporating the (implicit) effort decision which necessarily precedes any choice, we provide a tool for quantifying effort propensities across individuals and task designs. By accounting for endogenous effort decisions on tasks with varying levels of complexity, we are able to correct bias in estimates of risk aversion caused by mistakes due to inattention. By relating estimated effort

[1] We define cognitive noise as the residual randomness in individuals' choices, controlling for effort and in the absence of actual short-term preference fluctuations. Randomness in decisions plausibly has neural roots as decision values are formed from neural activity in the part of the brain called the ventromedial prefrontal cortex. The neural activity itself is stochastic (see Fehr and Rangel, 2011). Furthermore, individuals may exhibit cognitive uncertainty (Enke and Graeber, 2023). They may be unsure of their true preference and randomize within an interval of uncertainty which depends on individual characteristics (Jagelka, 2024). In addition, the perception of task attributes may itself be noisy (e.g., Woodford, 2020), particularly for attributes which occur infrequently (see Frydman and Jin, 2022).

[2] Importantly, this type of bias can be expected to persist in repeated measurements and thus cannot be removed by applying standard techniques for dealing with measurement error such as the ORIV method popularized by Gillen, Snowberg, and Yariv (2019).

propensities from experimental tasks to real-world outcomes, we establish the external validity of our estimates and uncover a more general tendency at the individual level which extends across settings. Finally, by modeling the preference relation as expected utility maximization subject to cognitive noise in the form of an error shock, we lay the groundwork for the reconciliation of competing models of discrete choice.

Our analysis is in line with recent research in psychology and economics which recognizes that effort and imperfect perceptions influence observed measures even in controlled settings.[3] We apply ideas from the recent literature that links discrete choice models with concepts of Costly Reasoning (Alaoui and Penta, 2022), Rational Inattention (Matějka and McKay, 2015; Caplin and Dean, 2015; Caplin, Dean, and Leahy, 2022), Rational Imprecision (Steverson, Brandenburger, and Glimcher, 2019), Efficient Coding (Frydman and Jin, 2022), Cognitive Uncertainty (Enke and Graeber, 2023), Cognitive Imprecision (Khaw, Li, and Woodford, 2021), Imperfect Self-Knowledge (Jagelka, 2024; Dohmen and Jagelka, 2024), or Limited Attention (Barseghyan, Molinari, and Thirkettle, 2021). As such, this paper enriches the broader domain of behavioral inattention summarized by Gabaix (2019).

A key innovation is that we separate the part of randomness in observed decisions which responds to the costs and benefits of answering a particular task according to one's latent preference (effort) from the part which is person-specific but invariant to task characteristics (cognitive noise).[4] A micro-founded modeling of the effort decision enables us obtain a more accurate picture both of preferences, and of apparent preference instability (cognitive noise), which we show to be less prevalent than indicated by previous research.

In our framework, a person first decides whether a task is worth paying attention to. Rather than assuming an exogenous mistake probability, we allow this decision to depend on the perceived costs and benefits of exerting effort to make a choice.[5] Our model therefore treats the reliability of each decision as the result of an endogenous process. If the perceived benefits of effort exceed the costs, the individual subsequently chooses the option which provides the highest expected utility, given task attributes, a latent preference of interest, and an error shock reflecting cognitive noise. Otherwise, the individual makes an effortless choice, which does not require the evaluation of

[3]This is evidenced by frequent inconsistent choices on repeated tasks in experiments (e.g., Hey and Orme, 1994; Gaudecker, Soest, and Wengstrom, 2011; Choi et al., 2014; Beauchamp, Cesarini, and Johannesson, 2017; Bruner, 2017; Gillen, Snowberg, and Yariv, 2019; Nielsen and Rehbeck, 2022) and by test-retest correlations well below the noise-free benchmark of "1" for repeated survey measurements elicited on the same sample within a short enough time period (e.g., a few weeks) such that the underlying attributes of interest can reasonably be assumed stable (e.g., Krueger and Schkade, 2008; Soto and John, 2017; Falk, Neuber, and Strack, 2021; Dohmen and Jagelka, 2024).

[4]The answer reliability measure of Dohmen and Jagelka (2024) and the cognitive uncertainty (CU) measure of Enke and Graeber (2023) manifestly contain various mixtures of these two distinct sources of randomness. For example, Enke and Graeber (2023) state that CU is "a composite measure that potentially captures people's awareness of a multitude of cognitive imperfections" and that "participants are relatively consistent in their degree of CU in a given domain", which suggests it largely captures what we call cognitive noise. However, they also find that their CU measure has some responsiveness to task complexity, which is a shifter in the cost of effort required to answer according to one's latent preference.

[5]For a theoretical analysis of conditions under which reasoning can be modeled as a cost-benefit analysis, see Alaoui and Penta (2022). The authors find that these conditions are weak.

expected utilities of the proposed choice alternatives, and answers randomly.[6]

In order to grasp the intuition behind our estimation strategy, it is useful to use an analogy with standard factor analysis methods. These extract a latent factor (here risk preference) from a large measurement system in which each observed measure can load differently on the latent factor. Our approach is similar in that observed differences in choice inconsistency are used to form effort probabilities which act as choice specific weights. These are then used to distinguish between choices deemed informative of structural parameters (risk aversion) and choices that are subject to randomness and thereby less informative of risk aversion.

We estimate the model on a representative sample of 1,224 Canadian high school seniors, each of whom made choices on 55 incentivized tasks used to elicit risk preferences.[7] There are two types of such choice tasks in this experiment. Both use the Multiple Price List (MPL) setup, which relies on ordered groups of binary choice tasks between lotteries with different expected payoffs and payoff variances, but differ in the complexity of those tasks.[8] The simpler design is based on tasks employed by Holt and Laury (2002) while the more complex design is inspired by tasks used by Eckel and Grossman (2008).[9]

Within each MPL of the simpler design, the first and the last task entails choices which should be easy for most individuals. In addition, there is a clearly visible pattern in the changing attractiveness of the riskier lottery. This reduces the per-task cognitive load necessary to make a choice according to an individual's latent risk preference compared to the more complex design which lacks these features. One might thus expect more mistakes and more noise on the more complex design due to (rational) inattention. We quantify this intuition.

We find that mistakes due to low effort increase with task complexity, with low relative stakes, and with fatigue—instances in which the costs of making choice in line with one's underlying preferences are higher and the benefits are lower. Changing the task design from the more complex one to the simpler one results in a 30% increase in the likelihood of exerting sufficient effort for the median individual. While 75% of the cross-sectional variation in individual choices on the simpler tasks is explained by a single variable—whether an individual's coefficient of relative risk aversion lies above or below the theoretical threshold at which a person should be indifferent between a given pair of lotteries—this accounts for only 20% of the cross-sectional variation in choices on tasks of the more complex design. Accordingly, underlying risk preference accounts for 90% of the explained cross-sectional variation in an individual's average choices on tasks of the simpler design but only 50% on the more complex tasks (the other half is noise due to inattention).

[6] A complementary approach in the recent literature assumes that individuals are more likely to take mental shortcuts when a setting is more familiar (see Cerigioni, 2021; Frydman and Jin, 2022).

[7] Several recent papers analyze aspects of this rich dataset (e.g., Belzil, Maurel, and Sidibé, 2021; Jagelka, 2024).

[8] Ordering ensures that the relative attractiveness of the riskier lottery is monotonically changing within an MPL.

[9] Harrison and Rutström (2008) provide an excellent summary on the various experimental designs and techniques used to elicit risk preferences in the laboratory. While multiple task designs exist, we lack a systematic understanding of the impact of design variations on decision noise and inferred risk preferences.

Accounting for endogenous effort is particularly crucial when observed choices contain a lot of noise. While the distribution of the coefficients of risk aversion based on the simpler tasks is largely unchanged if endogenous effort is omitted, omitting effort on the more complex design biases risk aversion estimates by approximately 50% for the median individual. We show that bias increases as an individual is more prone to errors, while the direction of the bias depends on an interaction between a particular task design and an individual's latent risk aversion. [10] This quantifies Andersson et al. (2016, 2020)'s claim that the interaction of random decision errors with an experimental design and an individual's latent risk preference may introduce bias in preference estimates, when sources of noise are not properly accounted for.

Our model has **high internal validity**. Estimated structural parameters explain 80% of the cross-sectional variation in the average number of risky choices and 70% of choices on any individual task. Structural estimates explain choices on the more complex tasks less well than on the simpler tasks, consistent with a bigger role of noise in decisions on the former.

Importantly, we also demonstrate **out-of-sample predictive power** which extends to a different decision context: choices between multiple lotteries. We find that (i) our risk aversion estimates from the binary choice tasks predict the coefficient of relative risk aversion implied by choices on the multiple choice tasks; (ii) our estimates of effort propensity predict the noisiness of a persons' decisions on the multiple choice tasks; and (iii) given the binary choice estimates of risk aversion and effort propensity, our model correctly predicts not only the direction of bias due to insufficient effort at the individual level but also explains much of its cross-sectional variation.

We show that our estimated effort propensity also has **external validity** and is particularly predictive of an individual's performance in low-stakes environments, notably on one of the most influential international assessment programs for mathematical literacy: the OECD-sponsored International Adult Literacy Survey score (IALS). We thus call it *low stakes motivation*. Transposing our result into the PISA international ranking which measures the same skills at age 15, we find that a one standard deviation increase in low-stakes motivation would affect the PISA numeracy ranking of a mid-performing country by approximately 9 places (out of 38). Furthermore, we provide evidence that the propensity to exert effort in low stakes settings may be fundamentally different from the propensity to exert effort in high stakes settings.

Even when individuals exert sufficient effort, residual randomness in choices from the point of view of the econometrician often remains (e.g., Dohmen and Jagelka, 2024). Such cognitive noise

[10] The relationship between bias and errors that we document complements and ties together evidence from the existing literature. For example, while Bruner (2017) claims that a *negative* relationship between mistakes and risk aversion is a general feature of monotone random choice models, Khaw, Li, and Woodford (2022) note that their "theory implies that increasing [the degree of imprecision] should both increase the randomness of the subject's choices and imply greater apparent risk-aversion" thus implying a *positive* relationship between mistakes and risk aversion. Cognitive uncertainty of Enke and Graeber (2023) also predicts bias in decisions between risky prospects (lower risk aversion for low payout probabilities and vice-versa). However, their proposed mechanism affects risky choice through probability weighting, which is a channel that cannot explain our results as about half of the tasks we use involve lotteries with a 50% probability of receiving either payment. In addition, we propose a strategy for *accounting* for the documented bias.

can be modeled as shocks to utility. A controversy arose recently in the literature as to where the stochastic shock should be placed when modeling economic preferences. In the standard additive Random Utility Model ("aRUM"), *the shock is appended to utility*. In a Random Preference Model ("RPM"), *the shock enters utility via preferences*. Apesteguia and Ballester (2018) prove that under standard assumptions on the utility function, the aRUM, unlike the RPM, exhibits anomalies in predicted choice probabilities under risk (and intertemporal delay) which call into question its continued use in preference estimation.

We demonstrate that estimated distributions of risk aversion using either aRUM or RPM shocks coincide once the decision to exert effort is incorporated. At least in the context of this experiment, proper estimation of the initial effort decision is empirically more important than the placement of the error term. Nevertheless, we use RPM shocks to preferences as our base specification due to their superior theoretical properties and to the intuitive interpretation of preference shocks as reflecting cognitive noise in the form of imperfect self-knowledge.

Existing estimates of the random preference model imply a significant degree of cognitive noise (a high estimated standard deviation of the preference shock). We show that after accounting for differences in endogenous effort, preferences are stable for the median individual. Furthermore, an individual's estimated degree of cognitive noise, unlike the propensity to pay attention, is independent of task design. This is what one would expect if the scale of a preference shock captures an individual characteristic such imperfect self-knowledge. Our findings thus complement Enke and Graeber (2023) and Enke, Graeber, and Oprea (2023), who find that inconsistencies in the domains of choice under risk, beliefs and expectations, and intertemporal choice are interrelated, Jagelka (2024) who shows that one personality trait—conscientiousness—predicts the stability of both risk and time preferences, and Dohmen and Jagelka (2024), who demonstrate that a single self-reported reliability measure predicts the test-retest consistency of survey measures of an individual's preferences, skills, and life satisfaction.

The rest of the paper is organized as follows: Section 2 surveys the literature on random choice models, Section 3 presents the structural model, Section 4 describes the data, Section 5 presents our estimates of the model parameters, Section 6 demonstrates out-of-sample predictive power and external validity of the estimates, Section 7 shows how our framework reconciles estimates from different discrete choice models, Section 8 discusses the broader implications of our findings for the design of preference elicitation tasks, and Section 9 concludes.

2 Background on Random Choice Models

The Random Utility Model (aRUM), which has its origins in Thurstone (1927) and Luce (1959), plays a central part in a multiplicity of microeconometric models of static and dynamic discrete choice. Its popularity has been stimulated by empirical research on consumers' discrete choices and by the development of the Conditional Logit model (McFadden, 1974). Although the aRUM may be used as a stochastic choice model, most applications incorporating an aRUM are concerned with deterministic choices. For instance, in the static discrete choice literature, the aRUM has been used as the main tool for specifying the demand for durable goods, in which the error term represents unobserved heterogeneity in tastes.

Because of its numerical simplicity, the aRUM model has been used extensively also in the experimental literature in which the cardinal utility shock reflects the degree of observed randomness in repeated choices which cannot be explained by variation in task parameters alone. The aRUM is used in many influential papers such as Hey and Orme (1994), Holt and Laury (2002), and Andersen et al. (2008). However, recent work by Wilcox (2011) and Apesteguia and Ballester (2018) point out that choice probabilities derived using the popular aRUM exhibit non-monotonicities which are at odds with a basic theoretical definition of risk (and time) preferences. For instance, the aRUM model predicts that individuals endowed with high risk aversion (for whom the utility function is very concave) would choose the safer and riskier options with equal probability.

Loomes and Sugden (1995) proposed the Random Preference Model (RPM) as a variant of random utility which adds an error term directly onto the coefficient of relative risk aversion, thus making it a random variable (or to an analogous parameter if another economic preference is studied). Apesteguia and Ballester (2018) prove that the RPM is monotone.

Although the RPM is intrinsically monotonic, it leaves no room for processing error. Unlike the aRUM, it cannot explain lapses in attention which may cause some individuals to choose dominated options.[11] The most common solution to this problem implemented in the experimental literature is to introduce a "tremble parameter" which captures the probability with which an individual makes a mistake (Harless and Camerer, 1994). In its original form, it essentially assumes that everyone evaluates the expected utility of each alternative and mistakes in decisions are purely random. The approach is still used, (see, e.g., Apesteguia and Ballester, 2018) who use a tremble parameter assumed to be common to the whole population.

Most efforts to relax this assumption have focused on modeling heterogeneity in the mistake probability, in general as a function of observed characteristics (see, e.g., Gaudecker, Soest, and Wengstrom, 2011; Andersson et al., 2020), while Jagelka (2024) also allows it to depend on unobserved heterogeneity. Such trembles – like exogenous additive random utility shocks – imply involuntary (exogenous) mistakes, even if propensity for them varies individual by individual. However, interpreting all mistakes as involuntary may be unrealistic. When individuals see the choice tasks as relatively complex or perceive little meaningful difference between the available choice options, they may judge that the disutility cost associated with introspection and solving the expected utility problem is too high compared with potential benefits of being able to make a choice in line with their latent preference. For this reason, we endogenize the decision to pay attention.

[11] In the RPM, the error term affects the preference parameter used to compare all alternatives. Therefore, no value of the shock can explain a choice which no level of risk aversion can justify.

Early attempts at incorporating the role effort into discrete choice models can be traced back to Hey (1995). Although no formal model of effort is presented, he operationalizes the intuition of Smith and Walker (1993) that "the error or randomness is determined optimally: the subject balances the gain from thinking about the question against the cost of so doing". Hey (1995) tests three potential parametrizations of the error shock variance, finding some support for the hypothesis that effort (proxied for by time spent on a task) reduces randomness. In a similar vein, Moffatt (2005), takes insights from the "capital-labour-production" framework of Camerer and Hogarth (1999) to introduce the possibility of learning (task order) into a tremble parameter. While he does outline a simple theoretical model of effort, instead of inferring it from observed choice patterns (like we do), he simply assumes it is measured by response time and does not take it into account when estimating risk preferences.[12]

3 Model

Before providing technical details, let us exposit the general set-up of the model: An individual makes choices on binary choice tasks designed to elicit a preference. Each choice provides information about the individual's latent preference of interest provided that he takes the task seriously.

When an individual is presented with a choice task, he first examines the readily and effortlessly available characteristics of the options among which he has to choose and decides whether or not it is worth to expend effort on making the choice. If it is, the individual exerts the amount of effort necessary to choose according to expected utility maximization given his relevant latent preference and cognitive noise represented by an error shock. If it is not, the individual answers according to a randomization strategy.

Consider a task involving a choice between two options: Y and X. An individual will choose Y when he prefers it and does not make a mistake or when he actually prefers X and makes a mistake because he previously decided the choice was not worth expending effort on and by chance picked the less preferred option.[13] We can write the probability that individual i *chooses* option Y on a binary choice task l as:

$$p(YC_{i,l} = 1) = p(E_{i,l} = 1) \cdot p(YP_{i,l} = 1) + \left[1 - p(E_{i,l} = 1)\right] \cdot p_{Y,i} \tag{1}$$

where $p(YC_{i,l} = 1)$ is the probability that individual i *chooses* option Y on choice task l; $p(YP_{i,l} = 1)$ is the probability that individual i *prefers* option Y on choice task l; $p(E_{i,l} = 1)$ is the probability that individual i will choose to exert effort on choice task l; and $p_{Y,i}$ is the probability with which

[12] A separate strand of the literature focuses on eliciting effort and cognitive noise through survey measures, without incorporating them into a formal random choice model (see, e.g., Enke and Graeber, 2023; Dohmen and Jagelka, 2024).

[13] One way of viewing our model, is as providing a micro-foundation for, and endogenizing, the popular "tremble" specification.

individual i picks option Y when he chooses not to exert effort. It can be understood as his "effortless" randomization strategy. A reasonable default value is $p_{Y,i} = 0.5$, i. e., in the absence of effort, an individual randomizes between the available options with equal probability.

We will now in turn characterize the initial effort decision and the determination of the *preferred* option given the relevant latent preference.

3.a Decision to Exert Effort

Each individual first briefly "takes in" a task, noticing its *readily and effortlessly available* characteristics. For the purposes of this exercise, we only consider such characteristics which pertain to the perceived costs and benefits of exerting *sufficient effort* to pick the preferred alternative on a given choice task.[14] Denote C_l the vector of readily and effortlessly available characteristics of choice task l which pertain to the perceived *costs* of exerting sufficient effort, and denote B_l the vector of readily and effortlessly available characteristics which pertain to the perceived *benefits* of exerting sufficient effort. Let us assume that the individual acts according to *net* perceived benefits and that effort is indivisible, i. e., conditional on choosing to exert effort, the individual will exert *sufficient* effort for making a choice according to his latent preference.[15]

Define an indicator, $E_{i,l}$, such that $E_{i,l} = 1$ when individual i decides to exert effort and $E_{i,l} = 0$ otherwise.The probability that individual i exerts effort when faced with choice l is given by:

$$\begin{aligned} p(E_{i,l} = 1) &= p(b_{0,i} + b_{1,i} \cdot B_l - b_{2,i} \cdot C_l + \varepsilon^b_{i,l} > 0) \\ &= p(\varepsilon^b_{i,l} > -b_{0,i} - b_{1,i} \cdot B_l + b_{2,i} \cdot C_l) \\ &= 1 - cdf(-b_{0,i} - b_{1,i} \cdot B_l + b_{2,i} \cdot C_l) \\ &= cdf(b_{0,i} + b_{1,i} \cdot B_l - b_{2,i} \cdot C_l) \end{aligned} \tag{2}$$

where $b_{0,i}$ is the intercept which captures individual differences in their baseline propensity to exert sufficient effort in the analyzed choice tasks (e.g., due to differences in personality or variability in how difficult the tasks are for different individuals), $b_{1,i}$ and $b_{2,i}$ are vectors of coefficients measuring the importance that individual i accords to each of the readily and effortlessly available characteristics pertaining, respectively, to the benefits and costs of effort. The model is completed by an error term $\varepsilon^b_{i,l}$ which reflects individuals' noisy perception of the actual costs and benefits of effort, something quite natural when making an (implicit) decision based on a quick preliminary look at a task. We assume that $\varepsilon^b_{i,l} \in (0, \infty)$ is an i.i.d. random shock and $cdf(\cdot)$

[14] Sufficient effort is the amount of effort which is necessary for an individual to be able to choose the alternative which yields higher expected utility given his latent preference.

[15] Modelling effort as a binary variable in a simple choice setting is consistent both with empirical evidence (e.g., Dohmen and Jagelka, 2024) and in line with existing theoretical models of information acquisition where the decision maker is often making a choice between paying a fixed information cost or not (e. g., Bartoš et al., 2016). Furthermore, adopting the sufficient effort framework allows us to abstract from defining the "units" of effort as well as from the fact that the same task may require a different amount of effort from different individuals. What matters from the point of view of the econometrician is not how effortful an individual found a task to be but whether or not he exerted sufficient effort to make a choice in line with his latent preference. Finally, while the effort decision is binary, the probability of exerting sufficient effort is continuous.

denotes the relevant cumulative distribution function given the distribution for $\varepsilon^b_{i,l}$.

3.b Preference Between Available Options

Assume that individual i is endowed with a utility function $U_i(\cdot)$ which maps a vector of attributes into utility. The attributes can be monetary values (m), non-pecuniary characteristics of interest (n), and other (nuisance) characteristics (o). Denote Ψ_i a vector of preference parameters over these attributes. In the presence of delay or intertemporal separation, discounted expected utility $DEU_i(m,n,o;\Psi_i)$ needs to be considered.

When an individual is faced with a choice between two options X and Y—in a deterministic world with perfect information on relevant attributes *and* conditional on exerting sufficient effort—he will prefer option Y if:

$$DEU_i(m_y,n_y,o_y;\Psi_i) > DEU_i(m_x,n_x,o_x;\Psi_i) \tag{3}$$

where m_y and m_x are monetary characteristics, n_y and n_x are non-pecuniary characteristics, and o_y and o_x are nuisance characteristics of options Y and X respectively.

However, for many individuals, observed choices reflect a degree of inconsistency which cannot be justified by variation in task characteristics alone. Besides insufficient effort, various forms of cognitive noise need to be considered (e.g., Loomes and Sugden, 1995; Kahneman, 2011; Enke and Graeber, 2023; Jagelka, 2024).[16] Indeed, even when individuals exert sufficient effort, residual randomness in individuals' choices from the point of view of the econometrician often remains, for example due to an individual's imperfect self-knowledge (e.g., Dohmen and Jagelka, 2024).

Cognitive noise can be incorporated by introducing shocks to utility: either additive shocks appended on to the utility function (leading to an additive random utility model or aRUM) or shocks directly affecting preference parameters (leading to a random preference model or RPM). We introduce a general error term ε_i to complete the model.[17] The discounted expected utility that an individual i derives from a choice option thus depends on choice characteristics, preferences, and shocks: $DEU_i(m,n,o;\Psi_i;\varepsilon_i)$. Certain contexts may favor one type of utility shock over the other. For example, Apesteguia and Ballester (2018) show that preference shocks have desirable theoretical properties when modeling risky choices.

When an individual is faced with a choice between two options in the presence of utility shocks, even conditional on exerting sufficient effort his preference over the options will be probabilistic unless one option is dominated by the other, i. e., there is no value of the error shock which would make it the preferred option. Without loss of generality, option Y is preferred when

[16] Alternatively, individuals may randomize deliberately, either because they have a *preference* for randomization (see Agranov and Ortoleva, 2017) or because randomization essentially allows them to achieve a lottery over available outcomes which they prefer to any individual outcome itself (see Cerreia-Vioglio and Riella, 2019).

[17] The subscript i reflects the fact that some individuals may be subject to less cognitive noise than others when making decisions, i.e., they receive smaller error shocks.

$DEU_i(m_y, n_y, o_y; \Psi_i; \varepsilon_i) > DEU_i(m_x, n_x, o_x; \Psi_i; \varepsilon_i)$.[18] The probability that individual i prefers option Y is therefore equivalent to the probability that the value of the shock is such that this inequality is satisfied.

To summarize: while utility differences (including error shocks) determine which option is preferred, the effort decision determines whether an individual converts the preference into an actual choice.

3.c Application to Risk Preference Elicitation

The general model is easily adapted to choice under risk: If sufficient effort is exerted, an individual will choose according to expected utility maximization given his coefficient of relative risk aversion and a *preference shock*, as in Jagelka (2024), i. e., a choice alternative is characterized by monetary attributes (payments and probabilities over them); the preference vector Ψ_i consists of the coefficient of relative risk aversion θ_i; the functional form for utility is constant relative risk aversion (CRRA); and the error shock ε_i is added directly on to the preference parameter. If sufficient effort is not exerted, the individual randomizes between the two options with equal probability, i. e., $p_{Y,i} = 0.5$.

Let $U_i(a)$ represent the utility which an individual obtains from a dollars. Define the coefficient of relative risk aversion $\theta_i = \frac{-a \cdot U''(a)}{U'(a)}$.[19] A CRRA utility function can then be written as:

$$U_i(a) = \frac{a^{(1-\theta_i)} - 1}{1 - \theta_i} = U(a, \theta_i) \tag{4}$$

We chose this representation of CRRA utility over the frequently used $U_i(a) = \frac{a^{(1-\theta_i)}}{1-\theta_i}$ (e.g., Andersen et al., 2008; Apesteguia and Ballester, 2018) due to its smoother convergence to $ln(a)$ in the immediate vicinity of $\theta = 1$. For a lottery X with two possible outcomes, x_1 dollars with probability p_{x_1} and x_2 dollars with probability $1 - p_{x_1}$, an individual's expected utility is: If $\theta_i \neq 1$,

$$EU_i(X) = p_{x_1} \cdot \frac{x_1^{(1-\theta_i)} - 1}{1 - \theta_i} + (1 - p_{x_1}) \cdot \frac{x_2^{(1-\theta_i)} - 1}{1 - \theta_i} \tag{5}$$

If $\theta_i = 1$,

$$EU_i(X) = p_{x_1} \cdot ln(x_1) + (1 - p_{x_1}) \cdot ln(x_2) \tag{6}$$

When making a choice between lottery X and lottery Y, an individual first receives a realization of a preference shock, ε_i. We assume that the shock affects the individual's *perception* of his latent risk preference embodied by the coefficient of relative risk aversion, θ_i, which represents the relevant coefficient of relative risk aversion that would prevail in a purely deterministic choice context.[20] The individual uses the shocked (or instantaneous) value of risk preference $\theta_i + \varepsilon_i$ to

[18] In full, option Y is preferred when $DEU_i(Y; \Psi_i; \varepsilon_{i,y}) > DEU_i(X; \Psi_i; \varepsilon_{i,x})$. When ε_i directly affects a preference parameter, $\varepsilon_{i,x} = \varepsilon_{i,y} = \varepsilon_i$ because both choice options are judged based on the same underlying preference. When ε_i is an additive utility shock, we can always combine the shocks to obtain $\varepsilon_i = \varepsilon_{i,y} - \varepsilon_{i,x}$ because *differences* in utility determine the preferred choice.

[19] We restrict θ_i to the (wide) range of risk aversion covered by the available elicitation tasks, so $\theta_i \in (-2, +5)$.

[20] For closed form solutions of the choice probabilities under the alternative random utility specification with addi-

compare the two alternatives. The expected utility of individual i from lottery X and lottery Y respectively becomes:

$$\begin{aligned} EU_i(X) &= p_{x_1} \cdot \frac{x_1^{1-(\theta_i+\varepsilon_i)}-1}{1-(\theta_i+\varepsilon_i)} + (1-p_{x_1}) \cdot \frac{x_2^{1-(\theta_i+\varepsilon_i)}-1}{1-(\theta_i+\varepsilon_i)} \\ &= EU(X;\theta_i+\varepsilon_i) \end{aligned} \tag{7}$$

and

$$\begin{aligned} EU_i(Y) &= p_{y_1} \cdot \frac{y_1^{1-(\theta_i+\varepsilon_i)}-1}{1-(\theta_i+\varepsilon_i)} + (1-p_{y_1}) \cdot \frac{y_2^{1-(\theta_i+\varepsilon_i)}-1}{1-(\theta_i+\varepsilon_i)} \\ &= EU(Y;\theta_i+\varepsilon_i) \end{aligned} \tag{8}$$

Assume that lottery X is less risky (has a lower variance in potential payoffs) than lottery Y in all lottery choice tasks l=1,...,L that an individual faces. The individual will prefer the riskier lottery Y to the safer lottery X on task l if

$$EU(Y_l;\theta_i+\varepsilon_{i,l}) > EU(X_l;\theta_i+\varepsilon_{i,l}) \tag{9}$$

The probability that Y is preferred on task l is equivalent to the probability that the value of the shock is such that the above inequality is satisfied. As $\varepsilon_{i,l}$ enters expected utility non-linearly, obtaining a closed-form expression for this probability is non-trivial. We follow Apesteguia and Ballester (2018) to do so, making use of the monotonicity of the random preference model (RPM).

Let us define a threshold level of indifference θ_l^{eq} which satisfies $EU(X_l,\theta_l^{eq}) = EU(Y_l,\theta_l^{eq})$, i. e., the level of θ at which any individual would be exactly indifferent between lotteries X and Y on choice task l in a deterministic context. We use the threshold level of indifference to obtain a closed-form expression for the probability that individual i prefers the riskier lottery Y on task l. Individual i will prefer the riskier lottery Y on task l if his shocked value of risk aversion is lower than the indifference threshold associated with task l:

$$\theta_i + \varepsilon_{i,l} < \theta_l^{eq} \tag{10}$$

or, rearranging, if the value of the shock is lower than $\bar{\varepsilon}_{i,l}$, the maximum value which still satisfies the inequality expressed in Equation (9):

$$\varepsilon_{i,l} < \bar{\varepsilon}_{i,l} = \theta_l^{eq} - \theta_i \tag{11}$$

Assuming that the random shock is normally distributed with $\varepsilon_{i,l} \sim N(0,\sigma_i^2)$, the probability that individual i prefers the riskier option Y on choice task l has a closed-form expression:[21]

$$p(YP_{i,l}=1) = \Phi\left(\frac{\theta_l^{eq}-\theta_i}{\sigma_i}\right) \tag{12}$$

The probability of preferring the safer option is simply:

$$p(YP_{i,l}=0) = 1 - p(YP_{i,l}=1) \tag{13}$$

Notice that an individual's risk preference can be understood as a normally distributed random

tive shocks (aRUM), please see Section B.a of the Online Appendix.

[21]Following Jagelka (2024), we restrict σ_i to plausible values, so $\sigma_i \in (0,1]$.

variable with mean θ_i and standard deviation σ_i, both of which are parameters to be estimated. We interpret θ_i as the individual's latent coefficient of relative risk aversion, which would prevail in a purely deterministic setting, and σ_i as a measure of either actual fluctuation in his risk preference or of the individual's degree of uncertainty as to its true value, i. e., as imperfect self-knowledge or cognitive noise. The lower an individual's σ_i, the more consistent is his risk preference over a panel of choices he has to make.

Both σ_i and $E_{i,l}$ impact the consistency of an individual's repeated observed choices. However, there is an important difference between the two. On the one hand, σ_i is related to the stability of preferences (or awareness of them). While those can vary somewhat from question to question, an individual would be choosing the expected utility maximizing option given his current (shocked) risk preference. On the other hand, by electing not to exert effort and instead choosing according to some heuristic he knowingly accepts the possibility of picking the *less preferred* option some percentage of the time. This would result in uninformative choices for the econometrician interested in inferring the individual's latent risk preference.

3.d Identification of Consistency Parameters

Both σ_i and $E_{i,l}$ measure the consistency of an individual's choice. However, each generates a specific pattern of choice inconsistency which allows for their separate identification.

3.d.i Identification Under Exogenous Effort

First, let us consider a simplified model in which an individual's decision to exert effort is insensitive to task-specific perceived costs and benefits of effort. In this case each individual would be characterized by a constant propensity to exert sufficient effort on all experimental tasks, $p(E_i = 1)$. If, in addition, the individual randomized with equal probability between the two options of a given task when he does not exert sufficient effort, he would make decision mistakes half of the time. Thus the individual would choose the option which gives him lower expected utility $\frac{1-p(E_i=1)}{2}\%$ of the time.

In this simplified case, identification is analogous to an RPM model with random trembles described in Jagelka (2024). We therefore only briefly outline the main intuitions here: In an RPM, no value of the preference shock can explain choices of dominated options. Several choice tasks in the present experiment involve such options and individuals choose them with non-zero probability. Only insufficient effort could explain such choices in our model and $p(E_i = 1)$ would therefore trivially be identified from such choices.

The constant effort propensity would be a source of uniform noise which affects all choices equally whereas σ_i, under a wide range of distributional assumptions on the preference shock, represents noise which has a higher chance to reverse a choice closer to an individual's point of indifference. It is identified from residual noise after stripping away the uniform noise component due to insufficient effort provision.

More generally, $p(E_i = 1)$ and σ_i can be identified from different moments of the noise distribution, even in the absence of dominated choices. Essentially, there is a tension between the occurrence of inconsistent choices on task with a θ_l^{eq} which is *close to*, or *far away from*, an individual's latent risk preference θ_i.[22] The resulting noise pattern is not sufficiently characterized by either consistency parameter alone.

3.d.ii Identification Under Endogenous Effort

Identification of endogenous effort parameters is more subtle than under exogenous effort, but follows the same general principles. The impact of shifters of the costs and benefits of effort is identified from systematic differences in noise patterns for tasks which they affect. For example, take two task designs eliciting the same latent preference but differing in complexity. Complexity is a shifter in the per-task cost of effort required for an individual to be able to choose according to his actual risk preference. If repeated choices on the more complex design are systematically more inconsistent/noisy than on the simpler design, the negative effect of complexity on effort would be manifested through the corresponding coefficient estimates in Equation 2.

Identification would break down if two task characteristics resulted in exactly the same noise pattern. Similarly, separate identification of the influence of a particular component of the effort decision from the preference shock would be compromised if that component resulted in an identical pattern of choice inconsistency as the preference shock, given the preference shock's assumed distribution. While unlikely in a sufficiently long panel of observed choices on tasks with enough variation in lottery characteristics (per individual, in a fixed effects estimation, or across individuals, in a representative agent framework), this should be evaluated on a case by case basis.[23]

3.e Estimation

Using Equation 1, an individual's contribution to the likelihood based on his choice on lottery choice task l is:

$$p(YC_{i,l} = yc_{i,l}) = p(YC_{i,l} = 1)^{YC_{i,l}} \cdot p(YC_{i,l} = 0)^{1-YC_{i,l}} \tag{14}$$

The likelihood contribution of individual i is the probability of jointly observing all L lottery choices he makes:

$$L_i = \prod_{l=1}^{L} p(YC_{i,l} = yc_{i,l}) \tag{15}$$

This is the likelihood to be maximized. We estimate the model individual by individual to obtain individual fixed effect estimates of the structural parameters.

[22]We define choice inconsistency as a deviation in choice from the one that would prevail in a purely deterministic setting given task parameters and the individual's relevant latent preference parameter.

[23]We verify this at the individual level by estimating our model with many random starting values and checking that the best fitting set of estimates is produced by a unique set of estimated structural parameters. For our base specification, this condition is satisfied approximately 99% of the time.

4 Data

We illustrate the usefulness of our model in improving estimates of risk preferences using experimental data from "The Millennium Foundation Field Experiment on Education Financing" designed by Claude Montmarquette and Cathleen Johnson.[24] This dataset fits our purposes for four main reasons: (1) it involves a large sample of 1,224 individuals, representative of the Canadian population on characteristics apart from age; (2) it features a long panel of 60 incentivized tasks per individual designed to elicit risk preferences; (3) while the elicitation tasks look similar, they include shifters for the costs and benefits of effort, e.g., they entail two levels of complexity; (4) each individual's performance on a low stakes and high-stakes test is recorded (an international numeracy test and high school GPA), which allows us to test the external validity of our estimates.

All 55 binary decision tasks involve choices between a safer and a riskier lottery.[25] They are organized into ordered groups (multiple price lists or "MPL") and displayed 5 at a time. Within each MPL, the relative attractiveness of the riskier lottery is either monotonically increasing or decreasing. Choice payments and probabilities are presented using an intuitive pie chart representation popularized by Hey and Orme (1994). Choices were incentivized and participants were paid for one randomly drawn decision at the end of the session. The availability of a long panel makes this an ideal setting to study decision noise at an individual level. Each choice provides information about an individual's risk aversion parameter provided that he takes the task seriously. The characteristics of the lotteries that are readily and effortlessly available to each individual, and therefore factor into the effort decision, are: task design and ordering (costs) and choice stakes (benefits).

Choice tasks of both the simpler (henceforth "sMPL") and more complex (henceforth "cMPL") type are designed to require little specialized skill, involve the same situation (pure choice under risk), and to be incentive-compatible (i.e., to provide an incentive for individuals to choose according to latent risk preference). "The key assumptions behind this set-up are that the individual understands probabilities and the expected values of options being offered, and that other factors that may affect risky choice besides latent preference (for example, wealth), can be controlled for adequately," (Dohmen et al., 2018). However, in reality these assumptions may not hold fully.

4.a Simple Multiple Price List (sMPL) Design

Of the 55 tasks designed to measure risk aversion, 30 are based on the work of Miller, Meyer, and Lanzetta (1969) and Holt and Laury (2002). There are three groups of ten questions each. In

[24] Participants were full time Canadian students in their last year of high school at the time of the experiment. The experiment was conducted using pen and paper choice booklets as well as simple random sampling devices like bingo balls and dice. Individuals were drawn from urban and rural schools in the provinces of Manitoba, Saskatchewan, Ontario and Quebec. See Table A.1 of the Appendix for descriptive statistics. For a full description, see Johnson and Montmarquette (2015).

[25] There are 5 additional multiple choice tasks, each of which involves a choice between 6 lotteries. We use them to test the out-of-sample performance of our model. They are described in more detail in Section 6.a c.

each group of questions, subjects are presented with an ordered array of binary lottery choices. In each choice task, they choose between lottery X (safer) and lottery Y (riskier). In each subsequent row, the probability of the higher payoff in both lotteries increases in increments of 0.1. For an example, see the left panel of Figure 1 below which shows the first two and last two choice tasks from ordered group of 10.

The sMPL design minimizes mental processing (effort) costs required to make a choice in line with one's latent risk preference. First, the initial choice in each ordered list of tasks is simple for most individuals as the safer lottery also offers a higher expected value. Second, the increasing attractiveness of the riskier option within each MPL is clearly visible due to the monotonically increasing probability of receiving the higher payment. Third, the last choice task is also simple as the higher payment is received with certainty and thus there is a dominated option.[26] This makes it a very simple and intuitive setting to elicit preferences.

4.b More Complex Multiple Price List (cMPL) Design

The remaining 25 tasks designed to measure risk aversion used in this study are a binarized version of the ordered lottery selection design developed by Binswanger (1980) and popularized by Eckel and Grossman (2002, 2008). A similar task design was used in Engle-Warnick, Laszlo, and Escobal (2006). They consist of five groups of five questions each. Once again, in each group of questions, subjects are presented with an ordered array of binary lottery choices. In each choice task, they choose between lottery X (safer) and lottery Y (riskier). This time, lottery X offers a certain amount in the first row and all other alternatives increase in expected payoffs but also in their variance. For an example, see the right panel of Figure 1.

While similar in appearance, the more complex "cMPL" task design lacks the three aforementioned features which reduce the per task effort required to make a choice in line with one's underlying risk preferences. We might thus expect choices to reflect a mix of signal from latent risk preference and noise due to inattention as more individuals may decide that the tasks are not worth the effort required to evaluate them correctly, given available incentives.

In a deterministic world, each individual should "switch" at most once between the riskier and safer option within an ordered group of tasks. Each person's "switching point" would then be indicative of their risk aversion. On the one hand, each individual should switch at exactly the same point on the 3 sets of sMPL questions.[27] On the other hand, under standard assumptions on the utility function (e.g., CRRA, CARA) the switching point should vary among the five sets of the cMPLs for a given individual even if he is paying full attention and consistently choosing according to his latent risk preference.[28] In a deterministic world, the sMPL tasks should allow

[26] In the last row of all three sets of sMPL questions designed to measure risk aversion, both lotteries offer the higher payment with certainty. Because no value of risk aversion can justify a preference for lottery X, it is dominated by lottery Y.

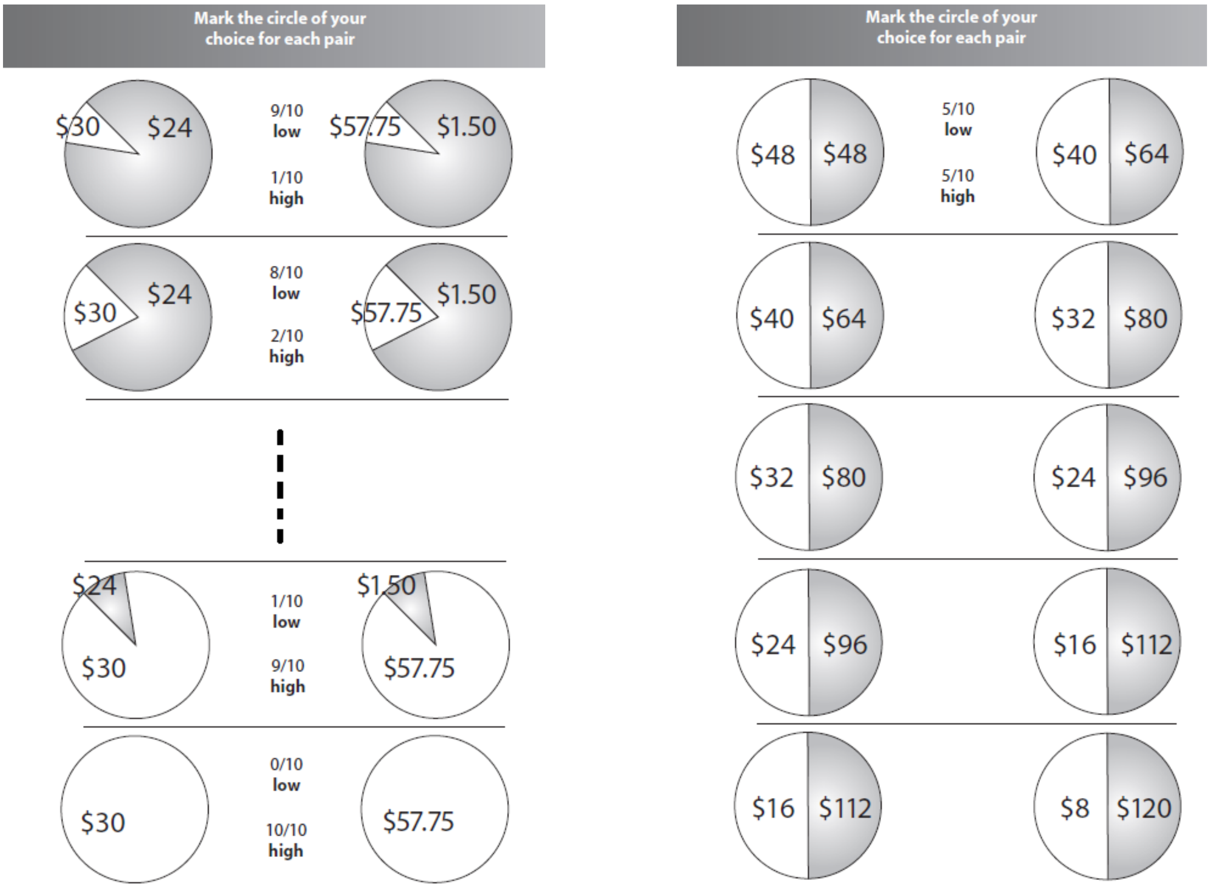
[27] This prediction holds for the popular constant relative risk aversion (CRRA) utility.

[28] Indifference thresholds for each of the 55 tasks in this experiment along with the percentage of individuals who

Figure 1: Binary Lottery Choice Tasks

(a) Simple "sMPL" Design

(b) Complex "cMPL" Design



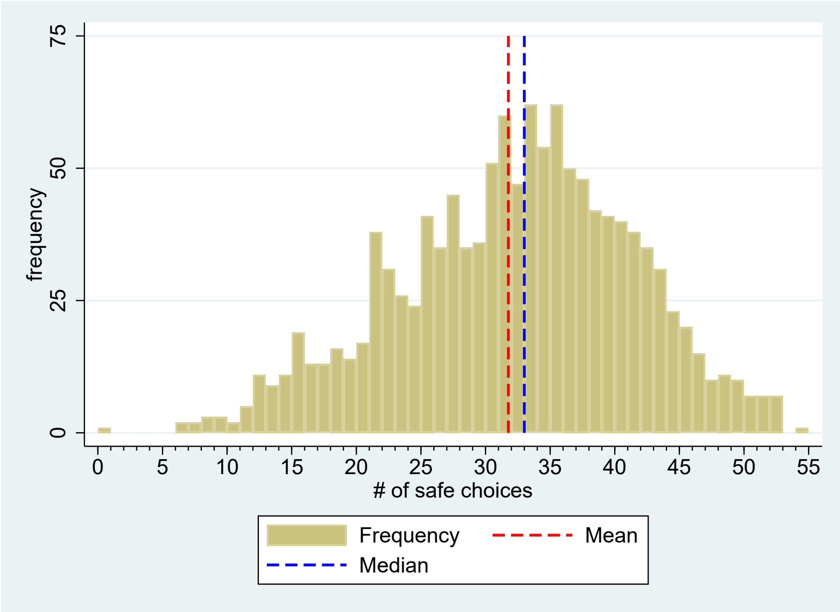
for the identification of an interval for an individual's risk aversion while the cMPL tasks should permit the refinement of this interval. Furthermore, while the sMPL tasks focus on the most common range of risk preferences (up to a coefficient of relative risk aversion of 1.37 under CRRA utility), cMPL tasks let us identify highly risk-averse individuals. The two types of task are thus complementary.

4.c Observed Individual Choices

Figure 2 plots the distributions of individuals' choices on tasks designed to elicit their risk preferences. Choices are heterogeneous and some individuals make decisions indicative of limit values of risk aversion - they either always choose the riskier or the safer lottery. The distribution of choices roughly resembles normality.

picked the riskier option on each task are displayed in Tables B.1 and B.2 of the Online Appendix. The three sets of choice tasks of the sMPL design share a common set of indifference thresholds under CRRA utility. The thresholds are increasing from Q1 to Q10 in each such MPL reflecting the increasing relative attractiveness of the riskier option. As predicted by the RPM model, the percentage of individuals choosing the riskier option is also monotonically increasing. The five sets of cMPL choice tasks are characterized by decreasing indifference thresholds which reflect a decreasing relative attractiveness of the riskier option. However, they do not exhibit the same congruence between the evolution of indifference thresholds and observed choices suggesting a more important role of noise on this task design and the need for a rich error specification in the structural model.

Figure 2: Distribution of Individual Choices on Lottery Tasks



Contrary to standard predictions, many individuals exhibit reversals in their choices within a given MPL.[29] This shows the usefulness of collecting data on the full set of tasks as opposed to assuming that each individual will maintain his choice after the "switching point" (as is often done in the literature, see Bruner (2017) for a recent example). In addition, some individuals also have inconsistent switching points across comparable MPLs. This is a more subtle form of choice inconsistency than outright reversals. If an individual is close to indifference around the switching point and he is somewhat uncertain as to his true preference, he may switch earlier on one set of tasks and later on another comparable set. While a small amount of cognitive noise may suffice to explain this behavior, choice reversals *within* a given MPL are indicative of highly erratic decision-making which suggests inattention.[30] These distinct patterns of choice inconsistency help separately identify the various parameters of the model which govern choice inconsistency, as discussed in Sections 3.d.ii and 5.a.iii.

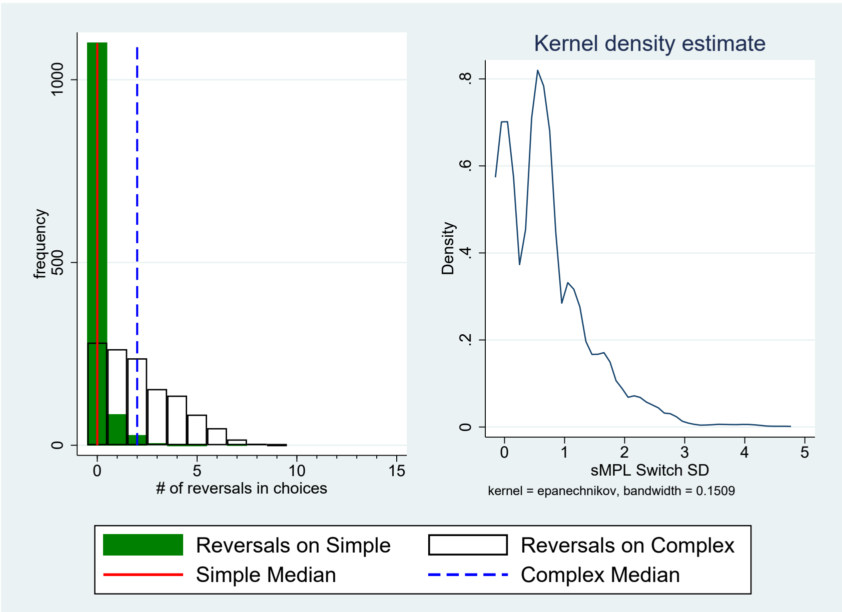
Figure 3 plots the distributions of reversals *within* a given MPL and of inconsistency in switching points *between* comparable MPLs. It reveals that while some reversals are observed on sMPL tasks, most of the action takes place on cMPL tasks. While almost 90% of individuals exhibit no reversal behavior on the former, 2/3 have apparent preference reversals on the latter. As mentioned above, while the sMPL design has features which minimize the per-task mental processing costs involved in choosing according to one's latent risk preference, making a choice according la-

[29]A reversal is defined as follows. Take for example one order list of the sMPL design which includes ten binary choice tasks ordered by increasing relative attractiveness of the riskier lottery. If an individual starts out by picking the safer option and then at some point switches to the riskier one as the riskier option becomes more attractive, this is considered standard behavior. If however he then reverts back to the safer option on the same set of tasks even though the riskier option became *even more* attractive, this is considered a reversal. The definition is analogous for lottery tasks of the cMPL design.

[30]Between choice tasks on a given MPL, there are fairly large jumps in the relative attractiveness of the riskier option.

tent risk preference on tasks of the cMPL design requires more mental effort. Hence we refer to the cMPL design as the more "complex" one. Some individuals may not find it worth their while to expend this effort and prefer to choose randomly at the cost of potentially choosing their less preferred option some of the time. This hypothesis is consistent with correlational evidence presented by Dave et al. (2010) who find that more complex risk elicitation tasks may lead to noisier behavior, especially in lower numeracy test subjects and with Jagelka (2024) who finds that variation in cognitive skills is the most important predictor of differences in individuals' propensity to make mistakes. It is supported by results from the structural model presented in the next section.

Figure 3: Observed Reversals per Individual on Lottery Choice Tasks



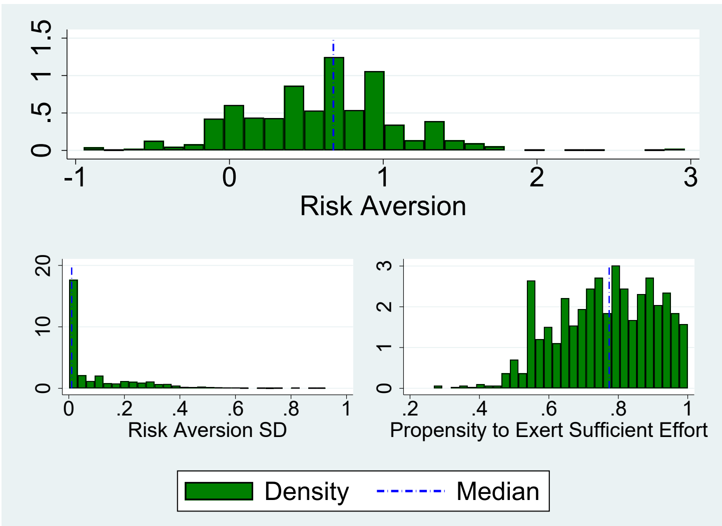
Inconsistencies in switching points can be easily detected on the three groups of sMPL tasks because they share common indifference thresholds under CRRA utility. We measure them as the standard deviation of switching points on the three ordered groups of the sMPL design for each individual (0 implies consistent switching points across the sMPL lists). The right graph of Figure 3 plots a distribution of switching point inconsistency on sMPL tasks smoothed through kernel density estimation. The sample distribution of inconsistent switching points looks similar to the sample distribution of choice reversals, with a high density at the origin and a fat tail. An analogous exercise cannot be done easily for the 5 groups of cMPL tasks as predicted switching points on them differ. Our structural model is needed to detect such inconsistencies.

The experiment also solicits background information collected both from students and from their parents. Descriptive statistics including demographic and socioeconomic variables for test subjects and their families are displayed in Table A.1.

5 Empirical Results

Estimates from the full model with endogenous effort and cognitive noise based on observed choices on all 55 lottery tasks show that the median individual is risk averse, exhibits almost no cognitive noise, and approximately 75% of the time exerts sufficient effort required for these tasks to give meaningful information about his latent risk preference. The median (mean) estimated values of the structural parameters are: 0.68 (0.88) for the coefficient of relative risk aversion, 0.01 (0.13) for the standard deviation of the coefficient of relative risk aversion (a proxy for cognitive noise or imperfect self-knowledge), and 0.77 (0.76) for the propensity to exert sufficient effort for choosing according to underlying preferences, averaged over the 55 tasks that each individual faced. Figure 4 plots the parameter distributions.[31]

Figure 4: Distributions of Structural Parameters Estimated Using the Model with Endogenous Effort



In order to put these results in context, it is helpful to compare them to existing estimates.[32] The obtained values of the coefficient of relative risk aversion are broadly in line with the previous literature (see e.g., Holt and Laury, 2002; Andersen et al., 2008; Apesteguia and Ballester, 2018; Apesteguia, Ballester, and Gutierrez, 2020; Jagelka, 2024).[33] While there are few existing estimates for the estimated scale parameter of the preference shock, previous results place it somewhere in the 0.3-0.6 range (see Apesteguia and Ballester, 2018; Apesteguia, Ballester, and Gutierrez, 2020; Jagelka, 2024), which is much higher than the value we obtain for the median

[31]The top histogram is capped at risk aversion of +3 as the overwhelming majority of observations falls within this range. There is a small spike again at +5, the highest level of risk aversion distinguishable with the available elicitation tasks.

[32]We omit our effort estimates from this discussion as we are not aware of any analogous previous estimates in the literature.

[33]While Holt and Laury (2002) do not report an estimate of the coefficient of relative risk aversion for the median individual, Table 3 of their paper implies that it is somewhere between 0.41 and 0.68 for the median individual on the "20x real" treatment, which most closely corresponds to the choice tasks included in this experiment. Andersson et al. (2020) obtain a lower estimate for the coefficient of relative risk aversion (0.25). However, the types of choice tasks that they use do not allow them to identify highly risk averse individuals.

individual.[34] We show that this discrepancy can be explained by the fact that when the initial effort decision is not taken into account, preference estimates based on the more complex choice tasks in our dataset are biased (see Section 5.a.iii). When using both simpler and more complex tasks in estimation, without taking into account how this difference in situations impacts effort decisions, the bias in estimates based on the more complex task design can be misinterpreted as preference instability or cognitive noise. We refer the reader to Section 5.b for a deeper discussion of this phenomenon.

We now describe in more detail the insights for theorists and practitioners revealed by our structural estimates.

5.a Endogenous Effort

Following our theoretical model, we allow the effort parameter to depend on readily and effortlessly available choice task characteristics. In the context of the lottery choices available in our dataset, these are: task design (complexity), task order (fatigue), and relative stakes (benefits of making the right choice).

The median individual is more likely to exert sufficient effort to choose according to latent preferences on less complex tasks, when stakes of getting the choice right are high, and when fatigue is low. The average impact of going from the more complex to the simpler task design is a 30% increase in the likelihood of exerting sufficient effort, p(E=1), for the median individual.[35] The marginal effect of increasing relative stakes by one standard deviation averaged across all 55 lottery choice tasks is a 7% increase in p(E=1) whereas increasing fatigue by one standard deviation results in a 2% decrease in p(E=1).[36]

Given the large estimated impact of experimental design on the cost of effort, we now explore its impact on the noise content of observed choices in more depth. To this end we first examine the predictive power of our structural parameters on moments of the raw data, and break it down by task design. This analysis clarifies the explanatory power of each structural parameter for the *average* behavior by an individual (both in terms of an average preference for the safer vs. riskier lottery and in terms of choice inconsistency) within a particular choice situation (task design). Second, we analyze the importance of the structural parameters in explaining *individual* choices. Third, we evaluate the bias in risk aversion estimates generated by omitting the initial endogenous effort decision and explain its determinants.

[34] The only estimate of a comparable magnitude comes from a sensitivity analysis from Apesteguia, Ballester, and Gutierrez (2020) using pooled individual estimates based on Coble and Lusk (2010) data and allowing for "correlation between parameters using a Gaussian copula".

[35] This is consistent with the pattern of choice inconsistency observed in the raw data, which is concentrated on the cMPL tasks (see Figure 3).

[36] We calculate these marginal effects using the estimated structural coefficients from our model. They are equal to the difference between an individual's predicted probability of exerting sufficient effort p(E=1) given each lottery's actual characteristics and the counterfactual *p(E=1)* if the design were flipped to cMPL, or if relative stakes or fatigue were increased by one standard deviation.

5.a.i Determinants of Average Behavior

We find that our model fits the data well. We take key moments of the distribution of individual choices and regress them on the estimated structural parameters: the preference parameter θ_i and consistency parameters σ_i and $p(E_i = 1)$.[37] Row 2 of Table 1 shows that these jointly explain over 80% of the cross-sectional variation in average choice behavior in terms of the percentage of the time that an individual selects the safer lottery and half of the variation in choice reversals. In comparison, the predictive power of demographic and socioeconomic variables is an order of magnitude smaller (see row 1 of Table 1).

Subsequent rows break down the explained variation in choices due to the estimated structural parameters into parts explained by the preference parameter and by the consistency parameters. This lets us compare their relative explanatory power, expressed as a percentage. Consistency parameters are further broken down into the standard deviation of risk aversion and the *propensity to exert effort*. This allows us to provide empirical evidence on the identification of the two types of consistency parameters based on different moments of choice inconsistency as outlined in Section 3.d.ii.

Almost 90% of the explained variation in observed choices is accounted for by the latent risk preference on the simpler choice tasks compared to only 50% on the more complex tasks (the remainder is noise due to inattention or imperfect self-knowledge). Increasing the coefficient of relative risk aversion by one standard deviation leads to a 15% increase in the proportion of safe choices selected on the simpler tasks, compared to a 10% increase on more complex tasks.[38] This is yet another indicator that actual risk preference has a larger impact on observed choices on the cognitively less demanding task design. It corroborates the large difference in noise content of the two task designs for eliciting risk preferences.

Cross-sectional variation in choice reversals - a strong form of choice inconsistency *within* an ordered group of tasks - is explained largely by differences in the propensity to exert sufficient effort on both task designs. This is consistent with the finding that the median individual exhibits stable risk preferences and choice inconsistency on lottery tasks is thus largely due to mistakes due to endogenous effort decisions.[39] However, cognitive noise, captured by the standard deviation of the coefficient of relative risk aversion, accounts for the majority of the explained cross-sectional variation in inconsistent switching points *between* groups of tasks in which a person with a given latent risk preference is predicted to switch at the same point, a more subtle form of choice inconsistency. While apparent preference instability and propensity to exert sufficient effort both explain randomness in observed decisions, they manifest through distinct patterns of choice in-

[37] We obtain an individual's propensity to exert effort $p(E_i = 1)$ as an average of the estimated task-specific effort propensities $p(E_{i,l} = 1)$.

[38] For more details, see Table B.3 of the Online Appendix which displays estimated regression coefficients along with calculated marginal effects.

[39] The standard deviation of risk aversion contributes 13% to the explained variation in reversals on the simpler tasks where individuals exhibit few reversals.

Table 1: Variation in Average Behavior on Lottery Choice Tasks Attributed to Preference vs. Consistency Parameters

		% Safe Choices	% Safe Choices: Simple	% Safe Choices: Complex	% Reversals	% Reversals: Simple	% Reversals: Complex	sMPL Switch SD
Demographic and Socioeconomic Variables	R2	0.05	0.04	0.07	0.02	0.03	0.02	0.03
All Parameters	R2	0.81	0.89	0.56	0.48	0.19	0.54	0.43
coefficient of relative risk Aversion		89.1%	88.2%	53.6%	0.0%	0.3%	0.0%	0.5%
Consistency Parameters		10.9%	11.8%	46.4%	100.0%	99.7%	100.0%	99.5%
- Stability		0.4%	0.8%	0.1%	0.0%	7.2%	0.1%	59.7%
- p(Effort)		10.5%	11.0%	46.4%	100.0%	92.5%	99.9%	39.8%

Notes: The rows labeled "R2" list the R2 of the regression of the moment listed in each column title alternatively on 18 demographic and socioeconomic variables and on the relevant estimated structural parameters of the model. Demographic variables include the student's sex, age, language, number of siblings living with him, his parents' age, as well as information on whether he was born in Canada and whether he is of aboriginal origin. These variables are available for 869 individuals. Socioeconomic variables include parents' level of education and income. The rows below represent the relative explanatory power of the relevant subgroups of parameters, expressed as a percentage. Columns 1-3 show the variation in the percentage of the time that a person chooses the safer option which is explained by observed characteristics and by the estimated structural parameters. Columns 4-6 show the explained variation in choice reversals. A reversal is defined as switching back to the safe option after having already picked the risky one on a given MPL even though the risky option became even more attractive, or vice versa. The last column looks at inconsistent switching points, a more subtle form of choice inconsistency. This analysis is only possible with tasks of the sMPL design which share a common set of indifference thresholds. The probability of exerting effort is averaged over the tasks of the relevant design (all; simple, i.e. sMPL design; complex, i.e. cMPL design) for each individual. The analysis excludes individuals with an estimated coefficient of relative risk aversion of below -2 and above +2 who are outside of the range of risk aversion captured by sMPL tasks. This leaves 1,109 observations or over 90% of the sample.

consistency and affect the two analyzed task designs to different degrees. These results illustrate the intuition behind the identification strategy outlined in Section 3.d.ii and complement the findings of Jagelka (2024).

Another interesting result is the lack of a relationship between the coefficient of relative risk aversion and choice reversals (see Table B.3 of the Online Appendix). This nuances Bruner (2017)'s claim that a negative relationship between mistakes and risk aversion is a general feature of monotone random choice models such as the RPM.[40]

5.a.ii Determinants of Individual Choices

We next examine how well our model predicts *each individual choice*. According to our model, an individual's choice on each lottery task is a function of the latent preference for risk only if the individual decides to exert sufficient effort. As discussed in Section 3, payoff-relevant lottery

[40] Bruner (2017) measured mistakes using choice tasks in which both alternatives have the same expected return and differ only in its variance (one option is thus stochastically dominated for individuals who are not risk neutral). In that situation, cognitive noise should in fact have a diminishing impact on observed choices for more risk averse individuals. However, this is a special case which applies to risk averse individuals on tasks with the same expected return where the threshold level of indifference is by definition 0—individuals with lower risk aversion than the threshold (who are risk-seeking) should choose the option with the higher variance while individuals with higher risk aversion (who are risk-averse) should choose the option with the lower variance. More risk averse individuals will have a coefficient of relative risk aversion further away from the threshold level of indifference and thus a given level of cognitive noise will be less likely to reverse their choice. There is no a priori reason to expect to see a negative relationship between risk aversion and choice inconsistency due to cognitive noise (let alone due to decision errors) on tasks where the threshold level of indifference varies such as the ones used in this experiment.).

characteristics (potential payoffs in the two lotteries between which an individual has to choose, along with their respective probabilities) can be conveniently summarized by a unique threshold level of risk aversion at which an individual would be indifferent between the two lotteries. Estimating a simple linear regression, Table 2 shows that, as implied by the model, an individual's coefficient of relative risk aversion being above or below the indifference threshold θ_l^{eq} for a given choice task (henceforth referred to as the "threshold dummy") is the most significant predictor of an observed choice on that task.[41] This information alone explains 75% of the cross-sectional variation in *individual* choices on lottery tasks of the simpler design. However, on tasks of the more complex design it explains only 21% of the cross-sectional variation in *individual* choices on lottery tasks. Once the threshold dummy is accounted for, the inclusion of the full set of payoff-relevant task parameters (lottery payoffs and their associated probabilities) in the regression has no meaningful impact. Adding an interaction between the effort parameter and the threshold dummy does not affect the ability of our model to predict choices on the simpler elicitation tasks but almost triples it for the more complex tasks.

Table 2: Explanatory Power of Individual Determinants of Lottery Choices

		Observed Choices			Wrong Choices		
		All	Simple	Complex	All	Simple	Complex
Demographic and Socioeconomic Variables	R2	0.00	0.00	0.01	0.00	0.00	0.00
Threshold Dummy	R2	0.46	0.75	0.21	0.01	0.00	0.00
p(Effort)	R2	0.00	0.00	0.01	0.24	0.16	0.18
p(Effort) * Threshold Dummy	R2	0.59	0.79	0.36	0.25	0.18	0.19
Full Set of Regressors	R2	0.62	0.82	0.40	0.28	0.25	0.22

Notes: The values displayed represent the R2 of a regression of observed individual choices (Columns 1-3) and of choices in which individuals did not select the expected utility-maximizing option (Columns 4-6) on various sets of regressors. Demographic and Socioeconomic Variables include the students' sex, age, language, number of siblings living with him, his parents' age, as well as information on whether the student was born in Canada and whether he is of aboriginal origin. Socioeconomic variables include parents' level of education and income. The „Threshold Dummy" is equal to one if the estimated coefficient of relative risk aversion is below the indifference threshold for a given task. „p(Effort)" is a task specific probability that an individual will exert sufficient effort given task characteristics and his estimated net benefit function. The Full Set of Regressors includes demographic and socioeconomic variables, individual lottery choice task parameters, and all estimated structural parameters along with their interactions with the difference between each lottery's estimated threshold level of indifference and the estimated coefficient of relative risk aversion as well as with the „Threshold Dummy". The probability of exerting effort is averaged over the tasks of the relevant design (all; simple, i.e. sMPL design; complex, i.e. cMPL design) for each individual.

The last three columns of Table 2 show that the endogenous effort propensity (modeled as a function of relative stakes, task order, and task design) in and of itself accounts for virtually all of the explained variation in wrong choices observed in the experiment.[42] The threshold dummy and its interactions with the remaining structural parameters contribute minimally. Finally, it is noteworthy that the 18 included demographic and socioeconomic variables together predict neither observed nor wrong choices.

[41] The "threshold dummy" is equal to one if the estimated coefficient of relative risk aversion is below the indifference threshold θ_l^{eq} for a given task. In a deterministic world with full attention, this variable should explain *all* of the variation in observed choices.

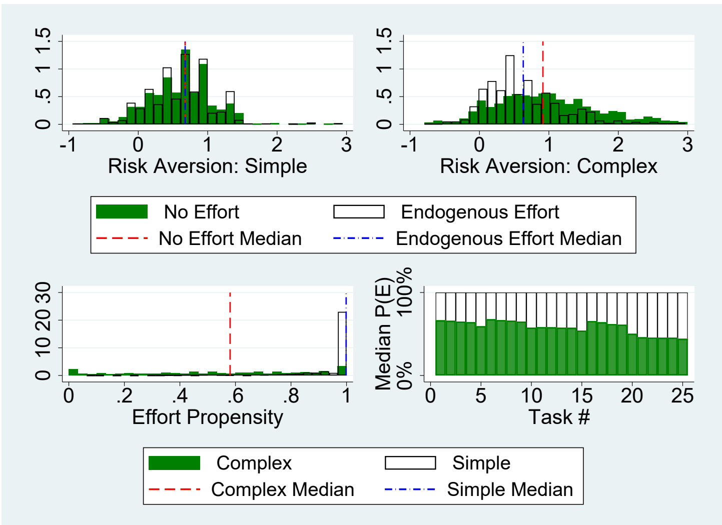
[42] "Ideal" choices are calculated for each choice task based on task parameters and each person's estimated latent risk preference. Wrong choices represent instances where the "ideal" choice differs from the observed one.

5.a.iii Task Design and Bias in Estimates

Having established that observed choices on one of the task designs in our experiment are a much noisier reflection of underlying risk preference than choices on the other task design, we now examine the consequences of this fact for preference estimates.

In the context of our experiment, relative stakes and fatigue only influence effort decisions on the more complex tasks (i. e., their estimated average marginal effect for the median individual on the simpler tasks is zero). This is easily discernible from the bottom right histogram of Figure 5, which plots estimated effort propensities for the median individual on the first 25 tasks of each design. Furthermore, when we average estimated effort propensities, for each individual alternatively across the 30 tasks of the simpler design and the 25 tasks of the more complex design (bottom left histogram in Figure 5), we find that the median individual exerts sufficient effort *all of the time* on the simpler tasks whereas the median individual only exerts it approximately 60% of the time on more complex tasks. This suggests that the available incentives are sufficient for the median individual on the simpler task design but not on the more complex one.[43] Accordingly, we find that while omitting the effort decision from our model leaves the distribution of estimated risk preferences from choices of the simpler design virtually the same (see top left histogram of Figure 5), doing so biases preference estimates from choices on the more complex design by approximately 50% for the median individual (see top right histogram of Figure 5).[44]

Figure 5: Distributions of Structural Parameters by Task Design



Andersson et al. (2016) conjecture that random decision errors will lead to an overestimation of risk aversion on lottery task designs in which individuals are expected to choose the riskier

[43]In contrast, the distributions of cognitive noise obtained using either task design are similar (see Figure 6). We discuss the implications of this finding in more detail in Section 5.b.

[44]The estimated coefficient of relative risk aversion using the more complex choice tasks is 0.6 when endogenous effort is accounted for and 0.91 when it is excluded. On the simpler tasks, the corresponding median is 0.68 *regardless* of whether the effort decision is estimated. As before, the histograms are capped at risk aversion of +3 as the overwhelming majority of observations falls within this range.

alternative more often than the safer one.[45] We test this hypothesis formally. For each individual, we first calculate the difference between the estimate obtained from the noisy complex design when the effort parameter is omitted and when it is included. This is the bias in risk aversion resulting from a naive model which does not take into account mistakes due to inattention. We next calculate the percentage of the time that the individual would be expected to choose the riskier option on the 25 tasks of the more complex design given his latent risk aversion. This represents the "lopsidedness" of this choice task design for each individual. The first column of Table 3 shows that bias is indeed increasing in the lopsidedness of the lottery choice tasks towards riskier choices.

Table 3: Bias as a Function of Individual's Predicted Percentage of Risky Choices and Choice Inconsistency on cMPL Tasks

	Estimated Bias in CRRA coefficient of relative risk Aversion	
Variables	(1)	(2)
Predicted % Riskier	2.70***	0.33***
	(0.82)	(0.093)
p(No Effort)		-0.14
		(0.09)
Predicted % Riskier * p(No Effort)		7.62***
		(0.23)
Constant	-0.11***	-0.024
	(0.027)	(0.033)
Observations	1,224	1,224
R-squared	0.472	0.722

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$

Notes: Each column presents the results of a regression of the estimated bias in the coefficient of relative risk aversion on variables predicted determinants of this bias. The "Predicted % Riskier" variable is the percentage of the studied binary choice tasks on which an individual would be predicted to choose the riskier lottery, given the tasks parameters and our estimate of that individual's latent coefficient of relative risk aversion $\hat{\theta}_i$. The "p(No Effort)" variable is our estimate of the percentage of the time that a given individual will choose randomly on the available decision tasks, i.e., the percentage of the tasks on which he will choose not to exert sufficient effort to make a choice according to his latent risk preference.

The bias should be larger for individuals who are less likely to exert sufficient effort on the choice tasks and are thus more prone to making mistakes. In the second column we add the estimated probability of not exerting effort along with the interaction term. The interaction term is significant and positive as predicted. Bias is highest for individuals who are prone to mistakes when their actual risk preference would lead them to disproportionately choose the risky lotteries in choice tasks they face. The marginal effect of increasing the predicted percentage of riskier choices by one standard deviation is a 0.88 increase in the bias of the estimated coefficient of relative risk aversion.[46] It can be understood as the effect of design imbalance at the individual level.

[45]When actual risk preference leads an individual to choose relatively many riskier options, random errors are more likely to flip the choice of a risky option to safe than the converse. This implies fewer observed risky choices than justified based on his latent risk preference and overestimation of risk aversion if decision error is not properly taken into account.

[46]The calculated marginal effect includes an interaction term calculated at the mean value of the estimated effort probability.

Given that that task complexity is the key determinant of endogenous effort in our setting, our findings predict a general relationship between elicitation task complexity and bias in preference estimates. As an illustration, consider a hypothetical set of multiple lists of tasks, each consisting of repeated binary choices eliciting the same parameter of interest (e.g., preference for risk, time, longevity) and assume that each list entails its own level of task complexity. Suppose that with full effort, each list should reveal the same decision pattern (the same sequence of choices). Now consider what happens if an individual reduces effort gradually when moving from the easiest to the most complex list. As the effort probability approaches 0, choices are made with an increasing degree of randomness until the probability of selecting each option reaches 0.5. Naive statistical inference which ignores the randomness in decisions will be biased as the observed choice pattern becomes disconnected from the one reflecting actual preferences. To take a concrete example, suppose we have a list with 10 decisions and assume that an individual has a level of risk aversion which leads him to chose 9 risky choices and 1 safe choice with full effort. Pure randomization (no effort) will result in a more balanced list of choices and will provide the false impression that the individual is more risk averse than they truly are (an upward bias). On the other hand, if the list is such that the individual prefers mostly the safe options, randomization will give the false impression that individuals are less risk averse then they truly are (a downward bias).

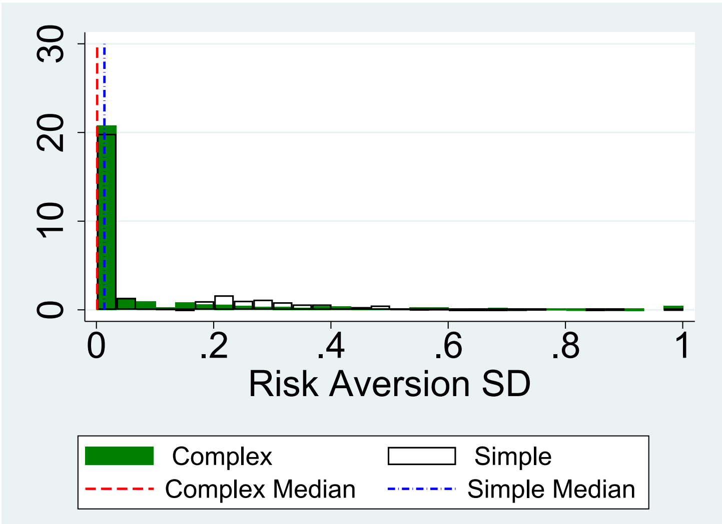
5.b Stability of Individuals' Preferences

A defining feature of the random preference model is that it assumes that the error term affects preference parameters directly, making them random variables. One possible interpretation is that each person has a "true" value of the preference parameter but some individuals have imperfect self knowledge and are essentially randomizing their choices within an interval around the true value. This is related to the concept of cognitive uncertainty examined by Enke and Graeber (2023). Another interpretation is that preferences do actually fluctuate due to external factors unobserved by the researcher such as fatigue or varying temperature in the room. It is one way of formalizing Kahneman (2011)'s observation that "[t]o a psychologist, it is self-evident that people are neither fully rational nor completely selfish, and that their tastes are anything but stable." Finally, individuals may randomize around their truly preferred choice because they actually have a *preference* for randomization (Agranov and Ortoleva, 2017).

The concept of unstable preferences is not standard in the economic literature and indeed there is a limit to how much preferences can plausibly fluctuate within a short time interval. Using the same dataset but estimating a model *without* endogenous effort, Jagelka (2024) shows that "for the median individual [by estimated risk aversion], choice inconsistency generated by the estimated preference shocks is concentrated within one or two cells from the switch point implied by constant preferences set at their average value". However, existing estimates of the scale of the error shocks are still large in absolute terms.

One of the contributions of this paper is to show that after accounting for differences in situations, preferences become stable for the median individual. A particular task design is a situation. Preference instability or cognitive noise estimated using only tasks of the same design is low. Furthermore, cognitive noise estimated separately on the simpler and more complex tasks is similar while the likelihood of exerting sufficient effort is different (see Figure 6 and the bottom left histogram of Figure 5, respectively). The fact that cognitive noise is the same across task designs while mistakes due to inattention vary suggests that the stability of preferences is an individual characteristic while decision errors are due to endogenous effort decisions, responsive to incentives.

Figure 6: Distributions of Estimated Cognitive Noise Parameter by Task Design



Once the decision to exert effort is incorporated into the model, the median individual has stable risk preferences even when all 55 available lottery choice tasks are used for estimation. Combined with the results from the previous section regarding bias arising from elevated noise on certain task designs, one may conclude that the high estimated standard deviation of risk preference shocks, when not accounting for differences in situations, is largely an artifact of biased preference estimates from tasks of the more complex design. This suggests that the failure to account for differences in situations results in an overestimation of preference instability or cognitive noise.

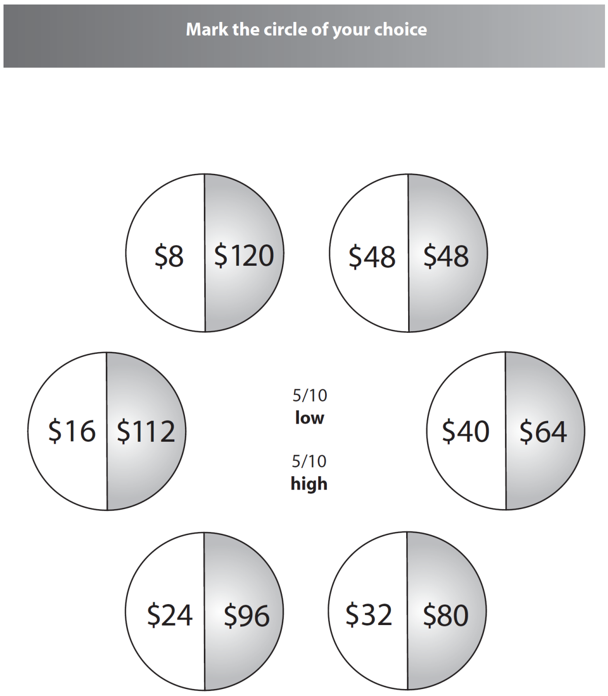
The inclusion of a properly parametrized effort parameter seems recommendable if one uses information on choices in different situations. We show that modeling inattention as a function of a few readily available attributes is able to account for differences in situations and greatly reduces the estimated degree of cognitive noise. Preferences nevertheless retain a degree of apparent instability for a fraction of the population. While the median individual has an estimated standard deviation of the coefficient of relative risk aversion of only 0.02, at the 75th percentile the standard deviation reaches 0.22 suggesting that there are individuals who are affected by significant cognitive noise, although they are in a minority. It is possible that once the influence of situations on choices is better understood, preferences will be revealed as essentially stable, in line with classical theory.

6 Out-of-Sample Predictive Power and External Validity

6.a Preference Elicitation Tasks with More than Two Options

In this subsection we test the ability of our estimates to predict behavior on a holdout sample of tasks involving many risky options. To this end, we make use of 5 observed choices, in each of which an individual can choose between 6 different lotteries. Each such multiple choice (MC) task combines the lotteries from an ordered group of 5 binary cMPL choice tasks into one task. See Figure 7 for an example combining the 5 cMPL tasks from Figure 1 into a single task.[47]

Figure 7: Lottery Choice Tasks - Multiple Choice cMPL design



We use the MC tasks to test the predictions of our model in a related but different setting. In this subsection we therefore do no further estimation. Instead, we take our estimates of individual effort propensities and risk aversion from the binary choice tasks, along with our structural model with endogenous effort, to the additional multiple choice data.

Our model implies a general relationship between elicitation task complexity and bias in inferred preferences (risk, time, social, etc.) when the relationship between task complexity and individuals' propensity to exert effort is not accounted for. Given our estimate that the median individual does not exert sufficient effort to make a choice in line with his latent risk preference almost a third of the time on the more complex binary choice tasks, we expect the level of risk aversion implied by individuals' decisions on the *even more complex* multiple choice data to also be bi-

[47]The 5 resulting MC tasks are analogous to the design used by Eckel and Grossman (2002).

ased. We derive a simple formula to predict this bias at the individual level and demonstrate its usefulness on the holdout sample data.

We proceed in three steps:

1. We first calculate the *naive* coefficient of relative risk aversion $\theta^N_{i,m}$ implied by an individual's observed choice on each of the five MC tasks m. The calculated risk aversion indifference thresholds, at which an individual would be indifferent between two lotteries, for the 5 ordered groups of binary cMPL choice tasks (see Table B.2 of the Appendix) give the respective thresholds for the 5 MC tasks.[48] As we perform no estimation here, we simply take the average of the 2 adjacent indifference thresholds to obtain the relevant $\theta^N_{i,m}$.[49] We then calculate the implied coefficient of relative risk aversion $\theta^N_{i,MC}$ that would be inferred from all 5 individual i's actual multiple choice decisions jointly:
$$\theta^N_{i,MC} = \frac{\sum_{m=1}^{M} \theta^N_{i,m}}{M} \tag{16}$$
According to our model, it will be biased for individuals who do not put in sufficient effort to make a choice in line with their latent risk preference.

2. For each of the five MC tasks, we then determine an individual's *preferred* lottery based on our estimate of that individual's latent $\hat{\theta}_i$, obtained by applying our full model to all 55 binary lottery choice tasks.[50] We calculate the implied coefficient of relative risk aversion $\theta^P_{i,m}$ that would be inferred from an individual's choice of his preferred option on multiple choice task m. To this end, we use the same indifference threshold methodology described in Step 1 above. We then obtain the implied coefficient of relative risk aversion $\theta^P_{i,MC}$ that would be inferred if individual i chose his preferred option on all 5 MC tasks, by averaging the constituent $\theta^P_{i,m}$, analogously to Equation 16.

3. Finally, for each individual i and MC task m, we calculate the predicted level of bias in the naive $\theta^N_{i,m}$ implied by the individual's choice on task m, given individual i's estimated latent risk aversion $\hat{\theta}_i$ and his estimated propensity to exert sufficient effort on the more complex

[48]The indifference thresholds now represent the level of risk aversion at which an individual would be indifferent between two *adjacent* lotteries in a given MC task. Individuals with a θ_i above the highest indifference threshold in a given MC task will prefer the safe lottery. Individuals with a θ_i below the lowest indifference threshold in a given MC task will prefer the riskiest lottery. Individuals with intermediate θ_i will prefer one of the remaining 4 lotteries, depending on their exact level of risk aversion. This holds under the simplifying assumption of fully stable/known risk preferences. It is supported by our finding that once effort is taken into account, the scale of the error shock tends towards 0 for the median individual.

[49]If an individual chooses either the safe lottery or the riskiest lottery, we only have one indifference threshold to work with. We thus either add half of the average difference between two adjacent indifference thresholds in the corresponding row of Table B.2 of the Appendix (if an individual chose the safe lottery on a given MC task) or subtract it (if he chose the riskiest lottery).

[50]As this section tests the out-of-sample predictive of our model, we need to distinguish between *estimates* of the coefficient of relative risk aversion $\hat{\theta}_i$ – obtained through maximum likelihood by applying our model to individual i's observed binary choices – and values of the coefficient of relative risk aversion $\theta^N_{i,m}$ implied by the individual's multiple choice data and calculated independently without the use of any statistics or econometrics. Our model will have out-of-sample predictive power if $\hat{\theta}_i$ predicts $\theta^N_{i,m}$.

binary tasks, $\hat{P}(E_{i,cMPL} = 1)$.[51] We obtain it as the difference between the biased $\theta^B_{i,m}$ that our model *predicts* to be implied by individual *i*'s choice under insufficient effort, and $\theta^P_{i,m}$ which would have been obtained from the individual's choice of his truly preferred option under sufficient effort.[52] According to our model summarized in Equation 1, $\theta^B_{i,m}$ will be a weighted average between $\theta^P_{i,m}$ (chosen when the individual exerts sufficient effort, so $\hat{P}(E_{i,cMPL} = 1)$ percent of the time) and the average coefficient of relative risk aversion that would be inferred from a random choice among the available options (when the individual chooses not to exert sufficient effort). More precisely, our model predicts that the bias $B^M_{i,m}$ in the inferred coefficient of relative risk aversion, for individual i based on his choice on MC task m with z options is:

$$B^M_{i,m} = \theta^B_{i,m} - \theta^P_{i,m} \tag{17}$$

with

$$\theta^B_{i,m} = p(E_{i,m} = 1) \cdot \theta^P_{i,m} + [1 - p(E_{i,m} = 1)] \cdot \frac{\sum_{z=1}^{Z} \theta_{z,m}}{Z} \tag{18}$$

where $\theta_{z,m}$ is the coefficient of relative risk aversion that would be inferred from a choice of option z on multiple choice task m. We assume that individual i has a constant propensity to exert sufficient effort across the MC tasks, equal to his average estimated propensity to exert sufficient effort on the more complex cMPL binary choice tasks, so $p(E_{i,m} = 1) = \hat{P}(E_{i,cMPL} = 1)$. Equation 18 can easily be adapted to predict bias due to insufficient effort in other revealed preference elicitation settings (e.g., time preferences, social preferences) by substituting in the relevant preference level implied by an individual's choice of the various available options.

The bias correction is obtained by subtracting predicted bias implied by Equation 17 from the naive coefficient of relative risk aversion: de-biased $\theta_{i,m} = \theta^N_{i,m} - B^M_{i,m}$.

We are now ready to assess the out-of-sample predictive power of our risk aversion estimates based on observed binary choices between lotteries. We test three main hypotheses: (i) our estimate of an individual's latent coefficient of relative risk aversion $\hat{\theta}_i$ using binary choice tasks will predict the coefficient of relative risk aversion implied by his choices on the multiple choice tasks, *and* it will better predict the de-biased coefficient of relative risk aversion $\theta_{i,m}$ than the naive $\theta^N_{i,m}$ implied by the individual's choice on multiple choice task m; (ii) individuals for whom we estimated a lower propensity to exert sufficient effort on the more complex binary tasks will make more inconsistent choices on the multiple choice tasks; and (iii) predicted bias implied by

[51] $\hat{P}(E_{i,cMPL} = 1)$ can be seen as the *upper* bound on individual *i*'s propensity to exert sufficient effort on the MC tasks as these are even more complex than tasks of the binary cMPL design, while having on average the same level of stakes as the cMPL binary tasks and involving the same (or greater) level of mental fatigue, because the MC tasks come at the end of the choice task section. We thus take our predicted bias as a *lower* bound on actual bias. This hypothesis is supported by our empirical results presented below.

[52] We analogously define actual bias $B^A_{i,m}$ as the difference between the coefficient of relative risk aversion implied by individual *i*'s actual choice on MC task *m* and the individual's estimated latent coefficient of relative risk aversion $\hat{\theta}_i$.

Equation 17 will predict actual bias in the implied coefficient of relative risk aversion at the individual level *and* it will be a lower bound on actual bias as the multiple choice tasks are even more complex.

To test the first hypothesis, we alternatively regress the naive and de-biased coefficient of relative risk aversion implied by the multiple choice tasks on our estimate of each individual's latent risk aversion taking endogenous effort into account. Perfect predictive power of the estimated full model parameters on the multiple-choice behavior would imply a constant equal to 0 and an OLS coefficient of 1 on the full model parameter.[53]

Our results confirm that $\hat{\theta}_i$ estimated on binary choice tasks has predictive power out of sample. The fact that the estimated slope coefficient is less than 1 suggests that there is some attenuation in mapping the full-model estimates. This makes sense as the out of sample decisions involve a different context: multiple choice tasks. Furthermore, $\hat{\theta}_i$, which already accounts for potential bias in risk aversion estimates due to insufficient effort, better predicts the coefficient of relative risk aversion implied by choices on the MC tasks once we apply the simple bias correction implied by Equation 17. This holds for each of the 5 MC tasks in our dataset taken individually (see Table 4), and also when we consider an individual's choices on them jointly (see Table A.2 of the Appendix). Indeed, the average share of variation explained by $\hat{\theta}_i$ roughly doubles once we apply our bias correction, the estimated OLS coefficient increases and becomes closer to 1, and the estimated constant falls and becomes closer to 0.[54]

Table 4: Predictive Power of an Individual's Latent Risk Aversion for Explaining the Naive and De-Biased Coefficient of Relative Risk Aversion Implied by Choices on Each MC Task

	MC Decision 1		**MC Decision 2**		**MC Decision 3**		**MC Decision 4**		**MC Decision 5**	
	Naive $\theta^N_{i,m}$	De-Biased $\theta_{i,m}$	Naive $\theta^N_{i,m}$	De-Biased $\theta_{i,m}$	Naive $\theta^N_{i,m}$	De-Biased $\theta_{i,m}$	Naive $\theta^N_{i,m}$	De-Biased $\theta_{i,m}$	Naive $\theta^N_{i,m}$	De-Biased $\theta_{i,m}$
Full Model $\hat{\theta}_i$	0.34***	0.54***	0.45***	0.74***	0.15***	0.24***	0.47***	0.77***	0.18***	0.27***
	(0.02)	(0.03)	(0.04)	(0.04)	(0.01)	(0.01)	(0.04)	(0.04)	(0.02)	(0.02)
Constant	0.92***	0.61***	1.14***	0.40***	0.64***	0.66***	1.16***	0.51***	0.67***	0.67***
	(0.04)	(0.04)	(0.05)	(0.05)	(0.02)	(0.02)	(0.05)	(0.05)	(0.02)	(0.03)
Observations	1224	1224	1224	1224	1224	1224	1224	1224	1224	1224
R-squared	0.13	0.27	0.11	0.27	0.11	0.19	0.12	0.27	0.10	0.17

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$

Notes: Each column displays the results of a regression of the coefficient of relative risk aversion implied by an individual's choice on a lottery multiple choice task on our estimate of that individual's latent risk aversion $\hat{\theta}_i$. The full model $\hat{\theta}_i$ estimate is obtained by estimating the full model using individual i's choices on all 55 binary choice tasks. The naive $\theta^N_{i,m}$ is calculated based on individual i's choice on a given multiple choice task, using indifference thresholds associated with the constituent lotteries. The de-biased $\theta_{i,m}$ is obtained by applying the bias correction implied by Equation 17 to the naive $\theta^N_{i,m}$.

We next verify that individuals who have a lower estimated propensity to exert effort on cMPL tasks also have a higher dispersion in the naive coefficient of relative risk aversion implied

[53] Conversely, if the full-model coefficients were not predictive at all, the OLS coefficient should be zero and the constant would capture the average population parameter in the multiple-choice questions.

[54] The bottom panel of Table A.2 of the Appendix shows that the same does not hold if we use a *naive* estimate of risk aversion, omitting the first stage effort decision. Our model predicts that bias due to insufficient effort should be in the same direction on the binary cMPL tasks and on the multiple choice tasks. This makes the naive (biased) estimate of the coefficient of relative risk aversion a better predictor of risk aversion implied by choices on the MC tasks, but only before the bias correction is applied.

by their observed choices on the five MC tasks. This is the case as the correlation between $\hat{P}(E_{i,cMPL} = 1)$ and the standard deviation of the coefficient of relative risk aversion implied by individual i's respective choice on each of the five MC tasks is -0.21, statistically significant at the 1% level.

Finally, we test how well our model predicts bias in risk aversion that would be inferred from the MC tasks without taking endogenous effort decisions into account. We do so by regressing actual bias on the bias predicted by our model due to insufficient effort. While perfect fit with our third hypothesis still implies a constant equal to 0, this time we would expect a slope coefficient *greater than* 1 on predicted bias as it should be a lower bound on actual bias.[55] Table 5 reveals that our model is able to predict the bias in the naive $\theta^N_{i,m}$, which would be inferred from a person's choice on each individual MC task, as well as the bias in the average $\theta^N_{i,MC}$ that would inferred considering all 5 multiple choice decisions jointly (see column 6). The estimated constant is close to zero implying that there is little actual bias when our model predicts that there should not be any, particularly when we look at all five MC choices jointly. Furthermore, we cannot reject the hypothesis that the slope coefficient is greater than or equal to 1. Indeed, we estimate it at 1.6 when considering all 5 MC choices together.

Table 5: Actual vs. Predicted Bias in the Coefficient of Relative Risk Aversion Inferred from Choices on MC tasks without Taking Endogenous Effort Into Account

	MC Decision 1	MC Decision 2	MC Decision 3	MC Decision 4	MC Decision 5	**MC Average**
VARIABLES			**Actual Bias**			
Predicted Bias	1.24***	1.08***	1.44***	0.96***	1.37***	1.60***
	(0.08)	(0.07)	(0.12)	(0.07)	(0.12)	(0.08)
Constant	0.17***	0.13**	0.03	0.31***	0.06	0.04
	(0.04)	(0.06)	(0.03)	(0.05)	(0.03)	(0.03)
Observations	1,224	1,224	1,224	1,224	1,224	1,224
R-squared	0.17	0.16	0.10	0.12	0.10	0.27

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$

Notes: Each column displays the results of a regression of actual bias in an individual's coefficient of relative risk aversion on predicted bias, either implied by a given decision on a multiple choice task (Columns 1-5), or jointly by that individual's decision on all 5 MC tasks (Column 6). Actual bias $B^A_{i,m}$ is calculated as the difference between the naive $\theta^N_{i,m}$ implied by an individual's choice on a given multiple choice task (or, in the last column, as an average implied by his choices on all 5 multiple choice tasks) and that individual's estimated $\hat{\theta}_i$ using the full model based on all 55 binary choice tasks. Bias predicted by our model $B^M_{i,m}$ is calculated as the difference between the biased $\theta^B_{i,m}$ predicted by our model to be implied by individual i's choice under insufficient effort on a given multiple choice task (or, in the last column, as an average implied by his choice on all 5 multiple choice tasks) and $\theta^P_{i,m}$ that would be inferred if the individual put in sufficient effort to choose his preferred option on a given multiple-choice task.

Taken together, these results illustrate that our endogenous effort model generalizes to a setting with multiple choice options. Our estimates of individuals' risk aversion and propensity to exert sufficient effort predict out of sample behavior on choices between multiple lotteries. A simple formula derived from our theoretical model is effective in removing bias in risk aversion implied by observed choices without requiring any further estimation.

[55] Recall that the MC tasks are even more complex than the binary cMPL tasks.

6.b Estimated Effort Propensity as a Proxy for Low-Stakes Motivation

While the internal validity of our model is well documented, intriguing questions remain: (i) Does the estimated individual propensity to exert effort in a low-stakes experimental setting capture an individual's broader tendency to exert effort? (ii) If so, does it apply to low-stakes settings as well as to high-stakes settings?

To answer these questions, we need outcomes that involve similar individual characteristics but differ with respect to the incentives they provide (high stakes vs. low stakes). To achieve this, we make use of the pre-experiment survey which contains two different measures of student achievement: the International Adult Literacy Survey quantitative score (measuring an individual's numeracy skills) and high school GPA.

The Survey of Adult Skills (PIACC), which contains the IALS, is the most important International Large-Scale Assessment of adult skills. The test is regularly administered to representative samples of national populations and is meant to provide a basis for international comparisons of adult achievement. Like the more prominent PISA test, which is administered to individuals at the age of 15 only, it assesses both verbal and numeracy skills.[56] As documented in many OECD publications, both tests are meant to assess the capacity of individuals to use mathematical concepts in solving practical problems. Indeed, the first version of the PISA test was developed based on the IALS, which predates PISA (e.g., OECD, 2019).

However, large scale international achievement tests such as PISA and IALS tests have been criticized for several reasons, including the fact that they may be affected by non-cognitive dimensions such as effort which may distort international comparisons. This point is exemplified in Gneezy et al. (2019), who study the PISA exam and show that the effort-incentive gradient may vary substantially across countries.

In our experiment, the numeracy score, like other elements, is purely anonymous, and has no subsequent implications. This makes it a *low-stakes outcome*. Individual grades, on the other hand, are highly important for most students. High school grades have a huge impact on subsequent schooling choices and may even be used by potential employers as a screening tool. This makes it a *high-stakes outcome*.

To answer the first question, we regress IALS numeracy scores and high school GPA on the individual specific effort propensity, controlling for other skills, preferences, and characteristics. To answer the second question, we regress IALS numeracy scores and high school GPA on self-reported high school engagement.[57] To facilitate comparison, we standardize all variables apart from sex.

Table 6 shows that our individual-specific effort propensity estimates are predictive of observed

[56] Only the numeracy section of IALS was administered in the dataset we are studying.

[57] High school engagement is calculated based on self-reports (hours spent on homework, handing in homework on time, self-reported effort in high school).

outcomes. Effort estimated from low-stakes experimental tasks predicts both numeracy scores and high school GPA, even after controlling for self-reported skills, personality, and sex. It is a particularly good predictor of the low-stakes IALS outcome where it alone accounts for approximately 10% of the total explained variation after including all the aforementioned controls. Furthermore, low-stakes effort is a better predictor of the low-stakes outcome, while high school engagement is a better predictor of the high stakes outcome. In fact, our measure of high school engagement alone accounts for nearly 75% of the explained cross-sectional variation in high school GPA whereas it is statistically insignificant for IALS test scores, once self-reported math skills are included. The ratio of (i) explained variation in IALS scores accounted for by our low-stakes effort estimates and (ii) the explained variation in IALS scores accounted for by our high-stakes effort estimates, is 40 times higher than the same ratio calculated for high school GPA. Furthermore, the marginal effect of increasing low-stakes effort by one standard deviation is (much) higher for the low-stakes outcome while the marginal effect of increasing high-stakes effort by one standard deviation is (much) higher for the high-stakes outcome. The ratio of (i) the marginal effect of our estimated effort propensity on IALS scores and (ii) the marginal effect of high school engagement on IALS scores is more than ten times higher than the same ratio calculated for high school GPA.

Table 6: Predictive Power of Low-Stakes and High-Stakes Motivation on the IALS Achievement Test and High School GPA

VARIABLES	(1) IALS	(2) IALS	(3) HS GPA	(4) HS GPA
p(effort)	0.12***		0.09***	
	(0.03)		(0.02)	
HS Motivation		0.05		0.41***
		(0.03)		(0.03)
Cognitive Skills	x	x	x	x
Non-Cognitive Skills	x	x	x	x
Risk Preference	x	x	x	x
Sex	x	x	x	x
Constant	0.05	0.07	-0.15***	-0.05
	(0.04)	(0.04)	(0.04)	(0.04)
Observations	1,224	1,224	1,224	1,224
R-squared	0.19	0.18	0.29	0.38

Standard errors in parentheses.
*** p<0.01, ** p<0.05

Notes: All variables apart from sex are standardized to be mean 0 and standard deviation 1. Cognitive Skills include self-reported math, computer, problem-solving, reading, writing, and communication skills. Non-cognitive skills include proxies for emotional stability, extraversion, and conscientiousness. Risk preference is the coefficient of relative risk aversion estimated using the endogenous effort model based on all 55 binary lottery choice tasks.

These results are interesting for many reasons. First, they show that our estimate of low-stakes effort propensity has external validity. This suggests that we may be capturing a more general behavioral tendency, *low-stakes motivation*.[58] Second, they indicate a fundamental distinction between effort provided in low stakes environments and effort exerted in situations with poten-

[58] Even though both are designed to measure an individual's competencies, results from an achievement test and high school GPA differ in other ways than stakes. While we view stakes as the most salient difference relevant to our setting, part of our findings could be explained by these other differences.

tially large impacts on a person's future.[59]

The estimated marginal effect of low-stakes motivation on the numeracy test scores is meaningful in magnitude. Increasing effort by one standard deviation, holdings self-reported skills, personality, and sex constant, is predicted to increase an individual's numeracy score by 0.12 standard deviations. In order to provide an illustration of the implications of this result in terms of international comparisons, it is informative to make use of the proximity between PISA and IALS and extrapolate the estimated 0.12 standard deviation effect to a corresponding difference in international rankings. This is easy to do because results of the PISA test are standardized so that the mean score is 500, and the standard deviation is 100. A 0.12 standard deviation increase therefore corresponds to an increase of 12 points on the PISA test. We motivate our choice of the PISA test comparison by the fact that it is regularly administered to a larger and more stable set of countries than the IALS achievement test and frequently referenced in policy discussions (see, OECD, 2019).

If we take PISA numeracy results from 2009, the period when our experiment was conducted, for a middle of the pack country like Poland (rank 19 out of 38 studied OECD countries), this would be enough to move it up 7 places (to 12/38) while decreasing effort by one standard deviation would make it move down 11 places (to 30/38).[60]

Online Appendix Tables B.4 and B.5 provide additional interesting insights on the skills and preferences which impact numeracy achievement tests and high school GPA. For example, self-reported math skills are the single most important predictor of numeracy scores while conscientiousness is the single most important predictor of high-school GPA. However, the latter has almost no marginal explanatory power once high school engagement is accounted for. Furthermore, it is sufficient to control for self-reported math skill and our measure of high-stakes motivation loses all statistical significance in a regression on IALS numeracy scores.

7 Reconciliation of Competing Discrete Choice Models

In the traditional Random Utility Model with additive i.i.d shocks (aRUM), the error term is appended to an individual's utility. Apesteguia and Ballester (2018) show that the aRUM as traditionally specified is not monotone when applied to risk preferences. Intuitively, the likelihood of preferring the riskier option is not monotonic with respect to risk aversion under the aRUM because shocks are added onto the cardinal utility of each alternative. As risk aversion goes to infinity, the difference in cardinal utilities of any two payments goes to zero for standard util-

[59]The fact that low-stakes motivation retains some predictive power for high school GPA may be an artifact of the latter being a sum of many constituent task performances, some of which can be perceived as low-stakes. While our estimated low-stakes motivation (effort propensity) is a statistically significant predictor of high-stakes motivation (high school engagement), the correlation between the two is very low (<0.07).

[60]These results assume the same normalization of the obtained numeracy scores as is described by the OECD for their PISA methodology: we re-scale the scores such that they are mean=500, standard deviation=100. The distribution of scores in our sample resembles a normal distribution, in line with the official PISA description.

ity functions in which risk aversion is related to the curvature of utility (e.g., CRRA or CARA). Therefore, any additive shocks with a strictly positive scale parameter $\sigma_{\theta,i}^{RU}$ will at some point fully drive the decision maker's choice. The likelihood of preferring the riskier (and the safer) alternative will thus approach 0.5 in the limit.

Despite the non-monotonicity, both the CRRA coefficient of relative risk aversion θ and the error scale parameter σ are identified if we have multiple binary choices between lotteries with varying payments and payment probabilities for each individual. As we have such information, we can estimate the aRUM model even though we view the RPM as a theoretically more sound alternative. Given the prevalence of the aRUM in past structural research estimating risk (and time) preference due to certain attractive features (tractability and ability to explain choices of dominated options with one error shock), we consider it worthwhile to compare estimates using the two competing error specifications embedded within our endogenous effort framework and to examine whether the non-monotonicity problem of the aRUM retains empirical relevance once endogenous effort is incorporated. .

Jagelka (2024) finds that the aRUM-induced non-monotonicity in the probability of choosing the riskier of two options with rising risk aversion is empirically relevant in the context of the present dataset. Apesteguia and Ballester (2018) use Danish data from Andersen et al. (2008) to estimate both an aRUM and an RPM with trembles using a representative agent framework. They find that the RPM risk aversion estimate is 14% higher than that of the aRUM and that the difference increases for more risk averse subjects.

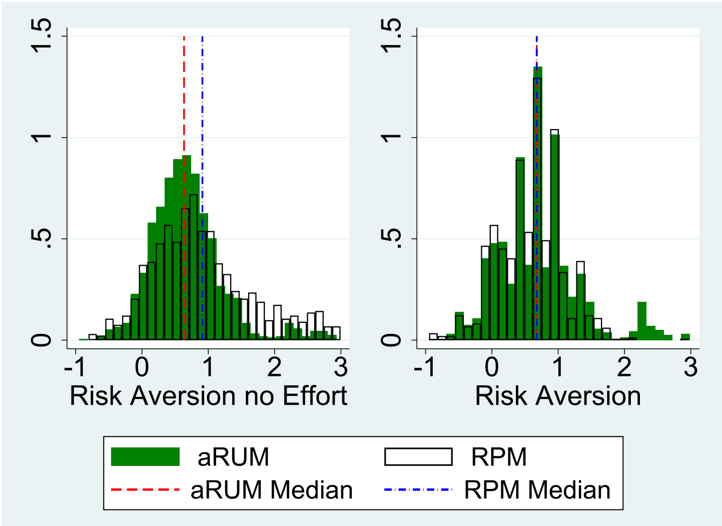
We corroborate these results when we estimate risk aversion *without taking into account the initial effort decision*. In this case, the entire distribution of the estimated coefficient of relative risk aversion is skewed to the right when using preferences shocks rather than additive utility shocks (see the left histogram of Figure 8 below).[61]

Once we estimate our model with endogenous effort, the non-monotonicity of the aRUM becomes empirically irrelevant, at least in the context of our experimental sample. The distributions of the coefficient of relative risk aversion estimated using either preference shocks or additive utility shocks converge (see the right histogram of Figure 8). Intuitively this is the case because after accounting for the endogenous effort decision, the estimated variance of the error shock falls both for the RPM and for the aRUM specification and approaches 0 for the median individual. While the predicted probability of choosing the riskier option under aRUM continues to be non-monotonic, the problematic behavior is shifted to high values of risk aversion which are not commonly observed.[62] After taking into account endogenous effort, one could thus put

[61] As before, the histograms are capped at risk aversion of +3 as the overwhelming majority of observations falls within this range.

[62] This is due to the fact that we are combining a non-monotone choice model (aRUM) with a quasi monotone one (random choice mistakes due to endogenous effort). Depending on the weight that each component receives, we can obtain a choice pattern which is more or less monotone. Given our empirical estimates of the structural parameters governing error shocks and endogenous effort, the non-monotone part receives little weight.

Figure 8: Distributions of Structural Parameters Estimated Using All Tasks with Alternatively the RPM and aRUM Error Structure



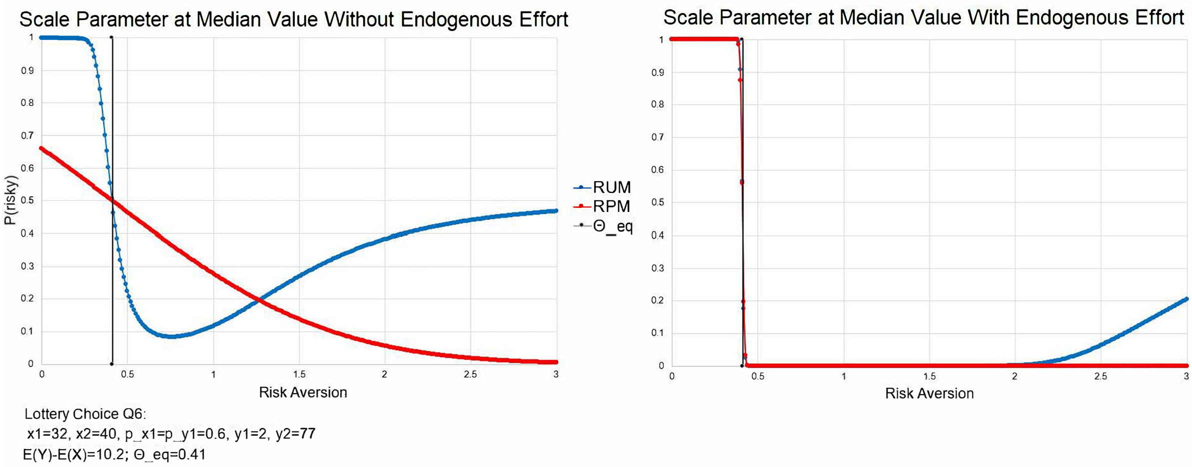
risk preference estimation within an aRUM framework in the same category as time preference estimation: theoretically problematic but empirically largely irrelevant.[63]

To illustrate this point, we take as an example the 6th choice task of the sMPL design contained in our data. In Figure 9 we plot the predicted probability of choosing the riskier lottery Y under RPM and under aRUM for values of risk aversion between 0 and 3 when the variance of the scale parameter σ_i is set at the median estimate using alternatively a model *without* endogenous effort (left) and our full model *with* endogenous effort (right). In either case, both the RPM and aRUM curves are initially decreasing, in line with the intuition that a more risk averse individual should be predicted to choose the riskier option with a lower probability. The curves cross at the threshold level of indifference for this choice task ($\theta_l^{eq} = 0.41$) where by definition the expected utilities of the two lotteries are equal and both models correctly predict that the probability of choosing either option is 0.5. The graph on the left assumes error shocks of a magnitude estimated for the median individual *when the effort decision is omitted*. The RPM curve continues to decrease monotonically while the aRUM curve reverts with risk aversion still below one (and thus while still at moderate and empirically frequent values of θ_i). It resembles Figure 1 in Apesteguia and Ballester (2018), which they use to illustrate the non-monotonicity problem of the aRUM. The graph on the right assumes error shocks of a magnitude estimated for the median individual *when the effort decision is endogenized*. Conditional on effort, the probability of choosing the riskier option becomes almost degenerate (deterministic). While it increases again for the aRUM, it does so at a much higher value of risk aversion. The non-monotonicity problem becomes practically irrelevant in terms of the empirical estimation of risk aversion using our data: the estimated distributions of the coefficient of relative risk aversion converge under RPM

[63]Apesteguia and Ballester (2018) also prove theoretical non-monotonicity when the aRUM is applied to the estimation of discount rates. However, they note that for standardly used experimental tasks the non-monotonicity occurs at "absurdly high" discount rates.

and aRUM once we allow the decision to exert effort to depend on an individual's perceived costs and benefits of doing so.

Figure 9: The RPM vs. aRUM Likelihood of Selecting the Riskier Lottery on the 6th Lottery Choice Task Assuming CRRA Utility



8 Implications for the Design of Preference Elicitation Tasks

Empiricists use a plethora of elicitation instruments for preferences, skills and other latent personal attributes. While these feature a number of design variations, these is a lack of a systematic understanding of their impact on the measurement properties of the chosen instrument. We study binary choices between safer and riskier lotteries of two designs—a simpler ("sMPL") design and a more complex ("cMPL") design—for eliciting risk preferences which were previously used interchangeably. On the one hand, we show that choices on tasks of the simpler design largely reflect an individual's latent risk preference. According to our estimates, 75% of the cross-sectional variation in individual choices on these tasks can be explained simply by whether an individual's coefficient of relative risk aversion lies above or below the theoretical threshold at which a person should be indifferent between a given pair of lotteries. The signal-to-noise ratio of observed choices is thus high and omitting either consistency parameter has little impact on the estimated distribution of risk aversion. On the other hand, our model with endogenous effort and cognitive noise reveals that only 20% of the cross-sectional variation in choices on *individual tasks* of the more complex design is explained by whether an individual's coefficient of relative risk aversion lies above or below the theoretical threshold. Furthermore, half of the explained cross-sectional variation in *average choices* on the more complex elicitation tasks can be attributed to random decision-making due to insufficient effort (in which case choices are uninformative about latent risk preference).

Omitting the initial effort decision results in estimates of risk aversion biased by 50% for the median individual on the more complex tasks. The bias is higher for individuals who have a high propensity to make mistakes and whose actual risk preference would disproportionately make them choose the riskier alternative. Our findings are in line with the predictions of our theoretical model which implies a general relationship between elicitation task complexity and bias in inferred preferences (e.g., risk, time, social). When endogenous effort is not accounted for, estimates are biased towards a preference level which would be consistent with a random choice pattern. We derive a simple formula which applied researchers can use to correct naive preference estimates. We demonstrate its effectiveness on a holdout sample with incentivized decision data from tasks involving choices between multiple lotteries.

Our results illustrate that a sophisticated error specification is much less important on tasks where individuals find it worthwhile to pay sufficient attention given the available incentives and choices are thus uncontaminated by decision error. It seems that the simpler choice design used in this experiment fits that description pretty well. Simple and complex models of behavior thus yield identical estimates of the population distribution of preferences. The inclusion of a properly parametrized effort parameter seems recommendable if one uses information on choices in different situations. A particular task design is a situation. At minimum, the noise content of a task design should be evaluated prior to proceeding with reduced form estimation.

Does this mean that sMPL tasks are better suited than cMPL tasks to elicit risk preferences and should thus be used exclusively? Not necessarily. In the context of the experimental dataset we examine, the two types of choice tasks are complementary. Assuming an appropriate econometric framework is used, researchers can employ them together to extract richer information on risk preferences. The calculated indifference thresholds displayed in Online Appendix Tables B.1 and B.2 illustrate that while the sMPL design covers the most common levels of risk-aversion, information from cMPL tasks can be used to narrow down the interval within which an individual's coefficient of relative risk aversion lies and to capture more extreme behavior at the high end of the distribution.[64] However, cMPL tasks will only provide valid preference estimates if choice inconsistency is properly accounted for. The sMPL design augmented to cover a wider range of risk preferences would seem recommendable, especially if reduced-form techniques are to be relied upon in estimation.

The obvious question is: What causes the large difference in individuals' effort decisions on the two task designs we study? As discussed in Section 4, the *ensemble* of features of the sMPL design work to minimize the per-task effort required to choose according to one's latent risk preference: the first and last choice in an ordered list are easy for most individuals and the progression in the relative attractiveness of the riskier lottery between them is clearly visible. This makes for a simple setting to elicit preferences, with low mental processing costs per choice and low cognitive demand. The amount of effort required to choose according to latent preferences on a given task, *sufficient effort* according to our definition, is thus sufficiently low such that most individuals find

[64] This is a feature of the particular parametrization of the sMPL tasks used in this experiment (which, however, is very standard in the literature, see e.g., Holt and Laury, 2002), rather than of the design itself.

it worthwhile given the experimental incentives.

The cMPL design lacks the aforementioned features which minimize the per task effort required to choose in line with one's actual risk preference. This makes the choices less intuitive and potentially requiring varying amounts of effort, depending on one's ease of processing the tasks which in turn likely depends on cognitive and non-cognitive skills. In this context, one can expect differentiation in the amount of mistakes made based on observed and unobserved heterogeneity. It is reflected in the wide dispersion of our estimated effort propensities on cMPL tasks.

One can conclude that while good experimental design can in some instances be used to substitute for modeling complexity, it is risky to rely on it alone. Even decisions on incentivized choice tasks in controlled experiments used to elicit a given preference reflect a mixture of signal and noise. The latter could become a strength once properly accounted for, as it can be used to understand the determinants of decisions not only when they go right (i. e., when they are consistent with a person's actual preferences) but also when they go wrong. This is particularly relevant in real-world settings which involve a high degree of complexity and choices likely contain a significant amount of noise. If we can identify factors which affect individuals' propensity to make mistakes in the laboratory, we might also be able to predict who and under what circumstances is prone to making sub-optimal decisions outside of it. This could in turn be used to design targeted interventions to help at risk individuals and thus contribute to redressing inequalities.

9 Conclusion

We develop and estimate a micro-founded random-choice model which accounts for endogenous effort and cognitive noise in estimates of preferences based on observed behavior. We exploit shifters of the costs and benefits of effort on choice tasks for eliciting preferences to demonstrate how our model can be used to (i) detect noise in observed choices, (ii) de-bias preference estimates, (iii) inform the policy implications of low stakes achievement tests such as PISA, and (iv) reconcile competing models of random choice.

Our model implies a general relationship between elicitation task complexity and bias in inferred preferences (risk, time, social, etc.) when the relationship between task complexity and individuals' propensity to exert effort is not accounted for. We apply it to experimental data from a representative sample of over 1,200 individuals, each of whom made 55 binary choices on incentivized tasks, commonly used to elicit risk preferences, of two designs which differ in their complexity. The availability of a long panel allows us to study preferences and decision noise at the individual level. When we omit the initial effort decision from the model, the estimated distribution of risk aversion based on the more complex choice tasks shifts, resulting in a bias of approximately 50% for the median individual.

We show that this bias arises from an interaction between the task design, an individual's latent risk preference, and his propensity to exert sufficient effort to make a choice in line with the latent preference. We use our model to derive a simple formula for the bias and demonstrate its predictive power oout of sample on incentivized decision data from tasks involving choices between multiple lotteries.

Individuals are less likely to exert the effort necessary to make a choice in line with their latent risk preference when mental processing costs and fatigue are high and when the stakes of making an incorrect choice are low. Unlike mistakes due to inattention, the estimated distribution of cognitive noise is invariant to elicitation task complexity. Indeed, preferences are stable for the median individual once effort is properly accounted for, suggesting that previous estimates misinterpreted differences in situation across decision tasks as cognitive noise. This is good news for traditional economic theory.

One of the advantages of having individual-specific estimates is that these may be used to test the external validity of the structural parameters of a model. We find that estimated effort propensity is predictive of an individual's performance in other low-stakes environments, even when controlling for measures of skills and demographics. This suggests that it captures a more general individual characteristic: low-stakes motivation. Extrapolating our results to contemporaneous PISA numeracy results, we show that a one standard deviation increase in low-stakes motivation would affect the international ranking of a mid-performing country by approximately 9 places (a 40% jump in the rankings).

Future applications of our model should aim to disentangle the impact of particular task design features on the noise content of observed choices. In addition, it is desirable to compare our method to reduced-form ways of detecting low quality responses such as asking individuals to self-report the overall reliability of their answers. The importance of low-stakes and high-stakes motivation in real-world settings also merits further study. Finally, the predictive power of economic preferences on outcomes should be re-evaluated once decision noise is accounted for and contrasted with the predictive power of the parameters governing the inconsistency of an individual's choices.

References

Agranov, Marina, and Pietro Ortoleva. 2017. "Stochastic Choice and Preferences for Randomization." *Journal of Political Economy* 125 (1): 40–68.

Alaoui, Larbi, and Antonio Penta. 2022. "Cost-Benefit Analysis in Reasoning." *Journal of Political Economy* 130 (4): 881–925. https://doi.org/10.1086/718378.

Andersen, S., G.W. Harrison, M.I. Lau, and E.E. Rutström. 2008. "Eliciting Risk and Time Preferences." *Econometrica* 76 (3): 583–618.

Andersson, O., H.J. Holm, J.R. Tyran, and E. Wengström. 2016. "Risk Aversion Relates to Cognitive Ability: Preferences or Noise?" *Journal of the European Economic Association* 14 (5): 1129–1154.

Andersson, O., H.J. Holm, J.R. Tyran, and E. Wengström. 2020. “Robust Inference in Risk Elicitation Tasks” [in En]. *Journal of Risk and Uncertainty* 61:195–209.

Apesteguia, J., and M.A. Ballester. 2018. “Monotone Stochastic Choice Models: The Case of Risk and Time Preferences” [in En]. *Journal of Political Economy* 126 (1): 74–106.

Apesteguia, J., M.A. Ballester, and A. Gutierrez. 2020. *Random Models for the Joint Treatment of Risk and Time Preferences* [in En]. Working Paper.

Barseghyan, Levon, Francesca Molinari, and Matthew Thirkettle. 2021. “Discrete Choice Under Risk with Limited Consideration.” *American Economic Review* 111 (6): 1972–2006.

Bartoš, Vojtěch, Michal Bauer, Julie Chytilová, and Filip Matějka. 2016. “Attention discrimination: Theory and field experiments with monitoring information acquisition.” *American Economic Review* 106 (6): 1437–1475.

Beauchamp, J.P., D. Cesarini, and M. Johannesson. 2017. “The Psychometric and Empirical Properties of Measures of Risk Preferences.” *Journal of Risk and Uncertainty* 54 (3): 203–237.

Belzil, C., A. Maurel, and M. Sidibé. 2021. “Estimating the Value of Higher Education Financial Aid: Evidence from a Field Experiment.” *Journal of Labor Economics* 39 (2): 361–395.

Binswanger, H.P. 1980. “Attitudes Toward Risk: Experimental Measurement in Rural India.” *American Journal of Agricultural Economics* 62 (3): 395–407.

Bruner, D.M. 2017. “Does Decision Error Decrease with Risk Aversion?” *Experimental Economics* 20 (1): 259–273.

Camerer, Colin F., and Robin M. Hogarth. 1999. “The effects of financial incentives in experiments: A review and capital-labor-production framework.” *Journal of Risk and Uncertainty* 19:7–42.

Caplin, Andrew, and Mark Dean. 2015. “Revealed Preference, Rational Inattention, and Costly Information Acquisition.” *American Economic Review* 105 (7): 2183–2203. https://doi.org/10.1257/aer.20140117.

Caplin, Andrew, Mark Dean, and John Leahy. 2022. “Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy.” *Journal of Political Economy* 130 (6): 1676–1715. https://doi.org/10.1086/719276.

Cerigioni, Francesco. 2021. “Dual decision processes: Retrieving preferences when some choices are automatic.” *Journal of Political Economy* 129 (6): 1667–1704.

Cerreia-Vioglio, David Dillenberger Pietro Ortoleva, Simone, and Gil Riella. 2019. “Deliberately Stochastic.” *American Economic Review* 109 (7): 2425–2445.

Choi, S., S. Kariv, W. Müller, and D. Silverman. 2014. “Who is (More) Rational?” *American Economic Review* 104 (6): 1518–50.

Coble, Keith H, and Jayson L Lusk. 2010. “At the Nexus of Risk and Time Preferences: An Experimental Investigation.” *Journal of Risk and Uncertainty* 41:67–79.

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

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2018. "On the Relationship Between Cognitive Ability and Risk Preference." *Journal of Economic Perspectives* 32 (2): 115–134.

Dohmen, Thomas, and Tomáš Jagelka. 2024. "Accounting for Individual-Specific Reliability of Self-Assessed Measures of Economic Preferences and Personality Traits." *Journal of Political Economy Microeconomics* 2 (3).

Eckel, C.C., and P.J. Grossman. 2008. "Forecasting Risk Attitudes: An Experimental Study Using Actual and Forecast Gamble Choices." *Journal of Economic Behavior & Organization* 68 (1): 1–17.

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

Engle-Warnick, Jim, Sonia Laszlo, and Javier Escobal. 2006. "The Effect of an Additional Alternative on Measured Risk Preferences in a Laboratory Experiment in Peru."

Enke, Benjamin, and Thomas Graeber. 2023. "Cognitive Uncertainty." *The Quarterly Journal of Economics* 138 (4): 2021–2067.

Enke, Benjamin, Thomas Graeber, and Ryan Oprea. 2023. "Complexity and Time." Technical report. National Bureau of Economic Research.

Falk, A., T. Neuber, and P. Strack. 2021. *Limited self-knowledge and survey response behavior.* Working Paper.

Fehr, E., and A. Rangel. 2011. "Neuroeconomic Foundations of Economic Choice–Recent Advances." *Journal of Economic Perspectives* 25 (4): 3–30.

Frydman, Cary, and Lawrence J Jin. 2022. "Efficient coding and risky choice." *The Quarterly Journal of Economics* 137 (1): 161–213.

Gabaix, Xavier. 2019. "Behavioral Inattention." In *Handbook of Behavioral Economics: Applications and Foundations 1,* 2:261–343. Elsevier.

Gabaix, Xavier, and Thomas Graeber. 2023. "The complexity of economic decisions." *Available at SSRN.*

Gaudecker, H.M., A. Soest, and E. Wengstrom. 2011. "Heterogeneity in Risky Choice Behavior in a Broad Population." *American Economic Review* 101 (2): 664–94.

Gillen, B., E. Snowberg, and L. Yariv. 2019. "Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study." *Journal of Political Economy* 127 (4): 1826–1863.

Gneezy, Uri, John A List, Jeffrey A Livingston, Xiangdong Qin, Sally Sadoff, and Yang Xu. 2019. "Measuring Success in Education: The Role of Effort on the test itself." *American Economic Review: Insights* 1 (3): 291–308.

Goossens, Jeroen, Arno M. Riedl, Bart Kuijpers, Eduard Ponds, Marike G. Knoef, and Sylvia J. Vos. 2023. "A comparison of risk preference elicitation methods and presentation formats." Forthcoming or update journal name if known, *To be updated with journal name, if applicable.*

Harless, David W, and Colin F Camerer. 1994. "The Predictive Utility of Generalized Expected Utility Theories." *Econometrica: Journal of the Econometric Society,* 1251–1289.

Harrison, G.W., and E. Rutström. 2008. "Risk Aversion in the Laboratory." In *Risk Aversion in Experiments,* 41–196. Emerald Group Publishing Limited.

Heckman, James J, Tomáš Jagelka, and Tim Kautz. 2021. "Some contributions of economics to the study of personality." In *Handbook of Personality: Theory and Research,* edited by O. P. John and R. W. Robins, 853–892. The Guilford Press.

Hey, J.D., and C. Orme. 1994. "Investigating Generalizations of Expected Utility Theory Using Experimental Data." *Econometrica: Journal of the Econometric Society,* 1291–1326.

Hey, John D. 1995. "Experimental investigations of errors in decision making under risk." *European Economic Review* 39 (3-4): 633–640.

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

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

Johnson, Cathleen, and Claude Montmarquette. 2015. "The lack of loan aversion among Canadian high school students." *Canadian Journal of Economics / Revue canadienne d'économique* 48 (2): 585–611.

Kahneman, D. 2011. *Thinking, Fast and Slow.* Macmillan.

Khaw, Mel W., Ziang Li, and Michael Woodford. 2021. "Cognitive Imprecision and Small-Stakes Risk Aversion." *The Review of Economic Studies* 88 (4): 1979–2013.

Khaw, Mel Win, Ziang Li, and Michael Woodford. 2022. "Cognitive Imprecision and Stake-Dependent Risk Attitudes." Technical report. National Bureau of Economic Research.

Krueger, Alan B., and David A. Schkade. 2008. "The Reliability of Subjective Well-Being Measures." *Journal of Public Economics* 92 (8-9): 1833–1845. https://doi.org/10.1016/j.jpubeco.2007.12.015.

Loomes, G., and R. Sugden. 1995. "Incorporating a Stochastic Element into Decision Theories." *European Economic Review* 39 (3-4): 641–648.

Luce, R.D. 1959. "Individual Choice Behavior: A Theoretical Analysis."

Matějka, Filip, and Alisdair McKay. 2015. "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model." *American Economic Review* 105 (1): 272–298. https://doi.org/10.1257/aer.20130047.

McFadden, D. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior in Zarembka." Academic Press, New-York, *Frontiers in Economics,* 105–142.

Miller, L., D.E. Meyer, and J.T. Lanzetta. 1969. "Choice Among Equal Expected Value Alternatives: Sequential Effects of Winning Probability Level on Risk Preferences." *Journal of Experimental Psychology* 79 (3p1): 419.

Moffatt, Peter G. 2005. "Stochastic choice and the allocation of cognitive effort." *Experimental Economics* 8:369–388.

Nielsen, Kirby, and John Rehbeck. 2022. "When Choices are Mistakes." *American Economic Review* 112 (7): 2237–2268.

Smith, Vernon L., and James M. Walker. 1993. "Monetary rewards and decision cost in experimental economics." *Economic Inquiry* 31 (2): 245–261.

Soto, C.J., and O.P. John. 2017. "The Next Big Five Inventory (BFI-2): Developing and Assessing a Hierarchical Model with 15 Facets to Enhance Bandwidth, Fidelity, and Predictive Power." *Journal of Personality and Social Psychology* 113 (1): 117.

Steverson, Kai, Adam Brandenburger, and Paul Glimcher. 2019. "Choice-Theoretic Foundations of the Divisive Normalization Model." *Journal of Economic Behavior & Organization* 164:148–165. https://doi.org/https://doi.org/10.1016/j.jebo.2019.05.026.

Thurstone, L.L. 1927. "A Law of Comparative Judgment." *Psychological Review* 34 (4): 273.

Wilcox, N.T. 2011. "Stochastically More Risk Averse:'A Contextual Theory of Stochastic Discrete Choice Under Risk." *Journal of Econometrics* 162 (1): 89–104.

Woodford, Michael. 2020. "Modeling Imprecision in Perception, Valuation, and Choice." *Annual Review of Economics* 12:579–601.

A Appendix

Table A.1: Sample Demographic and Socioeconomic Variables

Test Subjects	Observations	%	Mean	% if Male
Gender	1224			
Male		46%	NA	NA
Female		54%	NA	NA
Age	1224			
15-16		12%	NA	11%
17		67%	NA	65%
18		15%	NA	17%
19+		6%	NA	7%
Language	1224			
English		68%	NA	69%
Other		32%	NA	31%
Born in Canada	1087	96%	NA	96%
Lives with Siblings	1224	75%	NA	76%
Parents				
Age	1068	NA	46	NA
Indigenous Canadian	1224	7%	NA	7%
# Children under 18	1085	NA	2	NA
Thinks University is Important	1088	92%	NA	91%
High School Dropout	1224	12%	NA	11%
High School	1224	52%	NA	50%
University	1224	36%	NA	39%
Annual Income	976			
<20k		6%	NA	6%
20-40k		13%	NA	11%
40-60k		23%	NA	24%
60-80k		19%	NA	17%
80-100k		15%	NA	17%
100k+		24%	NA	25%

Table A.2: Predictive Power of an Individual's Risk Aversion Estimates to Explain the Naive and De-Biased Level of Risk Aversion Implied by Choices Averaged Across all 5 MC Tasks

	Multiple Choice Average	
	Implied Naive $\theta^N_{i,MC}$	De-Biased $\theta_{i,MC}$
Full Model $\hat{\theta}_i$ Estimate		
OLS Coefficient	0.32***	0.51***
	(0.02)	(0.02)
Constant	0.91***	0.57***
	(0.03)	(0.03)
Observations	1,224	1,224
R-squared	0.17	0.35
Naive $\hat{\theta}_{i,cMPL}$ Estimate		
OLS Coefficient	0.43***	0.46***
	(0.02)	(0.02)
Constant	0.74***	0.54***
	(0.03)	(0.04)
Observations	1,224	1,224
R-squared	0.28	0.25

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$

Notes: Each column displays the results of a regression of the coefficient of relative risk aversion jointly implied by an individual's choices on the 5 lottery multiple choice tasks on an estimate of that individual risk aversion from binary lottery choice tasks. The full model $\hat{\theta}_i$ estimate is obtained by estimating the full model using individual *i*'s choices on all 55 binary choice tasks. The naive $\hat{\theta}_{i,cMPL}$ estimate is obtained by estimating a model without endogenous effort using individual *i*'s choices on 25 binary choice tasks of the cMPL design. The naive $\theta^N_{i,MC}$ is calculated from individual *i*'s choices on 5 multiple choice tasks, using indifference thresholds associated with the constituent lotteries. The de-biased $\theta_{i,MC}$ is obtained by applying the bias correction implied by Equation 17 to the naive $\theta^N_{i,MC}$.

B For Online Publication

B.a aRUM Choice Probabilities

For a choice between lottery X and lottery Y under aRUM, we thus have:

$$\begin{aligned} EU_i^{RU}(X) &= p_{x_1} \cdot U_i(x_1) + (1 - p_{x_1}) \cdot U_i(x_2) + \varepsilon_{i,X}^{RU} \\ &= EU(X;\theta_i) + \varepsilon_{i,X}^{RU} \end{aligned} \tag{19}$$

and

$$\begin{aligned} EU_i^{RU}(Y) &= p_{y_1} \cdot U_i(y_1) + (1 - p_{y_1}) \cdot U_i(y_2) + \varepsilon_{i,Y}^{RU} \\ &= EU(Y;\theta_i) + \varepsilon_{i,Y}^{RU} \end{aligned} \tag{20}$$

Assuming that the two shocks are independent and normally distributed random variables, the probability that individual i prefers the riskier lottery Y on choice task l is:

$$\begin{aligned} p(risky)_{i,l}^{RU} = p(YP_{i,l} = 1)^{RU} &= P\left[EU_i^{RU}(Y) > EU_i^{RU}(X)\right] \\ &= P\left[\varepsilon_{i,Y}^{RU} - \varepsilon_{i,X}^{RU} > EU(X;\theta_i) - EU(Y;\theta_i)\right] \\ &= \Phi\left[\frac{EU(Y;\theta_i) - EU(X;\theta_i)}{\sigma_i^{RU}}\right] \end{aligned} \tag{21}$$

where $\varepsilon_{i,Y}^{RU} - \varepsilon_{i,X}^{RU} \sim N(0, \sigma_i^{RU^2})$ and $\sigma_i^{RU} \in (0, \infty)$.

B.b Calculated Indifference Thresholds

Table B.1: Indifference Thresholds and Observed Sample Proportions of Risky Choices on sMPL Type Choice Tasks

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
θ_{12}	-1.71	-0.95	-0.49	-0.14	0.15	0.41	0.68	0.97	1.37	Inf
% choosing risky sMPL Group 1	0.7%	0.9%	2.2%	8.5%	24.6%	38.2%	58.9%	79.2%	91.2%	99.8%
% choosing risky sMPL Group 2	0.3%	0.5%	1.2%	4.8%	15.6%	24.1%	43.1%	65.8%	85.9%	99.5%
% choosing risky sMPL Group 3	0.8%	0.9%	2.2%	6.1%	17.3%	26.8%	45.8%	68.3%	87.8%	99.4%

Table B.2: Indifference Thresholds and Observed Sample Proportions of Risky Choices on cMPL Choice Tasks

	Q1	Q2	Q3	Q4	Q5
θ_{12} cMPL Group 1	2.97	1.00	0.60	0.42	0.00
% choosing risky cMPL Group 1	70.5%	67.7%	53.7%	38.1%	34.9%
θ_{12} cMPL Group 2	4.73	1.69	1.06	0.78	0.00
% choosing risky cMPL Group 2	71.2%	72.8%	79.5%	65.3%	28.3%
θ_{12} cMPL Group 3	1.37	0.45	0.26	0.17	0.00
% choosing risky cMPL Group 3	48.7%	39.4%	30.3%	26.3%	14.4%
θ_{12} cMPL Group 4	4.46	1.50	0.94	0.68	0.00
% choosing risky cMPL Group 4	64.1%	79.8%	65.8%	45.8%	34.6%
θ_{12} cMPL Group 5	1.54	0.51	0.30	0.20	0.00
% choosing risky cMPL Group 5	41.3%	54.7%	45.3%	30.7%	19.5%

B.c Additional Results

Table B.3: Explaining Average Choices and Reversals on Lottery Choice Tasks Using Fixed Effects Estimates: Ordinary Least Squares Coefficients

	% Safe Choices	% Safe Choices: Simple	% Safe Choices: Complex	% Reversals	% Reversals: Simple	% Reversals: Complex	sMPL Switch SD
Risk Aversion	40.6***	34.3***	48.7***	-0.3	-0.1	-1*	0
	(0.82)	(0.41)	(1.63)	(0.32)	(0.12)	(0.55)	(0.04)
Risk Aversion SD	4.2**	0.2	10.1**	3.9***	1.7***	5***	2.7***
	(2.01)	(1.55)	(4.42)	(0.79)	(0.44)	(1.50)	(0.15)
Risk Aversion * SD	-9.8***	-8.7***	-9.6**	0.3	-0.4	2.1	-0.6***
	(1.99)	(1.55)	(4.41)	(0.78)	(0.44)	(1.49)	(0.15)
p(No Effort)	42.2***	35.3***	50.8***	18.1***	4.1***	19.3***	2.4***
	(1.02)	(0.73)	(1.01)	(0.40)	(0.21)	(0.34)	(0.07)
Risk Aversion * p(No Effort)	-55.7***	-58.2***	-60.0***	0.6	5.2***	-2.5***	0.7***
	(1.37)	(1.11)	(1.41)	(0.53)	(0.31)	(0.48)	(0.11)
Effect of Increasing Each Structural Parameter by One Standard Deviation							
- Risk Aversion	13.0	15.0	9.8	-0.1	0.1	-0.9	0.0
- Risk Aversion SD	-0.2	-0.8	0.7	0.6	0.2	1.0	0.4
- p(No Effort)	1.4	0.2	2.3	2.6	1.0	2.5	0.4

Standard errors in parentheses
*** p<0.01, ** p<0.05

Notes: The rows display the coefficient of the regression of the moment listed in each column title on the full set of structural parameter estimates, including interactions. Standard errors are in parentheses. A reversal is defined as switching back to the safe option after having already picked the risky one on a given MPL even though the risky option became even more attractive, or vice versa. The last column looks at inconsistent switching points, a more subtle form of choice inconsistency. This analysis is only possible with tasks of the sMPL design which share a common set of indifference thresholds. The probability of exerting effort is averaged over the tasks of the relevant design (all; simple, i.e. sMPL design; complex, i.e. cMPL design) for each individual. p(No Effort)=1−p(Effort). The effect of increasing each structural parameter by one standard deviation takes into account relevant interaction terms calculated at the average values of the variables each parameter is interacted with. The analysis excludes individuals with an estimated coefficient of relative risk aversion of below −2 and above +2. This leaves 1,124 individuals or over 90% of the sample.

Table B.4: Predictive Power of Low-Stakes and High-Stakes Motivation on the IALS Achievement Test

VARIABLES	(1) IALS	(2) IALS	(3) IALS	(4) IALS	(5) IALS	(6) IALS	(7) IALS	(8) IALS
p(effort)	0.15***	0.12***	0.14***	0.12***				
	(0.03)	(0.03)	(0.03)	(0.03)				
HS Motivation					0.11***	0.04	0.09**	0.05
					(0.03)	(0.03)	(0.03)	(0.03)
Math Skills		0.39***		0.36***		0.39***		0.36***
		(0.03)		(0.03)		(0.03)		(0.03)
Soft Skills				0.14***				0.13***
				(0.03)				(0.03)
Hard Skills				-0.00				-0.01
				(0.03)				(0.03)
Risk Preference				-0.02				-0.01
				(0.03)				(0.03)
Emotional Stability				-0.04				-0.04
				(0.03)				(0.03)
Extraversion				0.04				0.04
				(0.03)				(0.03)
Conscientiousness			0.09***	0.02			0.05	-0.00
			(0.03)	(0.03)			(0.03)	(0.03)
Sex				-0.09				-0.12**
				(0.06)				(0.06)
Constant	0.00	-0.00	-0.00	0.05	-0.00	-0.00	-0.00	0.07
	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)
Observations	1,224	1,224	1,224	1,224	1,224	1,224	1,224	1,224
R-squared	0.02	0.17	0.03	0.19	0.01	0.16	0.01	0.18

Standard errors in parentheses
*** p<0.01, ** p<0.05

Notes: All variables apart from sex are standardized to be mean 0 and standard deviation 1. Soft skills include self-reported reading, writing, and communication skills. Hard skills include self-reported computer and problem-solving skills. Risk preference is the coefficient of relative risk aversion estimated using the endogenous effort model based on all 55 lottery choice tasks.

Table B.5: Predictive Power of Low-Stakes and High-Stakes Motivation on High School GPA

VARIABLES	(1) HS GPA	(2) HS GPA	(3) HS GPA	(4) HS GPA	(5) HS GPA	(6) HS GPA	(7) HS GPA	(8) HS GPA
p(effort)	0.11***	0.09***	0.09***	0.09***				
	(0.03)	(0.03)	(0.03)	(0.02)				
HS Motivation					0.53***	0.49***	0.49***	0.41***
					(0.02)	(0.02)	(0.03)	(0.03)
Math Skills		0.30***		0.26***		0.22***		0.22***
		(0.03)		(0.03)		(0.02)		(0.02)
Soft Skills				0.25***				0.22***
				(0.03)				(0.03)
Hard Skills				-0.08***				-0.07***
				(0.03)				(0.03)
Risk Preference				-0.05**				-0.05**
				(0.02)				(0.02)
Emotional Stability				0.04				0.01
				(0.03)				(0.03)
Extraversion				-0.13***				-0.10***
				(0.03)				(0.02)
Conscientiousness			0.34***	0.22***			0.07**	0.03
			(0.03)	(0.03)			(0.03)	(0.03)
Sex				0.28***				0.10
				(0.05)				(0.05)
Constant	0.00	0.00	0.00	-0.15***	0.00	0.00	0.00	-0.05
	(0.03)	(0.03)	(0.03)	(0.04)	(0.02)	(0.02)	(0.02)	(0.04)
Observations	1,224	1,224	1,224	1,224	1,224	1,224	1,224	1,224
R-squared	0.01	0.10	0.13	0.29	0.28	0.33	0.29	0.38

Standard errors in parentheses
*** p<0.01, ** p<0.05

Notes: All variables apart from sex are standardized to be mean 0 and standard deviation 1. Soft skills include self-reported reading, writing, and communication skills. Hard skills include self-reported computer and problem-solving skills. Risk preference is the coefficient of relative risk aversion estimated using the endogenous effort model based on all 55 lottery choice tasks.