I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 18315

Separating Preferences from Endogenous Effort and Cognitive Noise in Observed Decisions

Christian Belzil
Tomáš Jagelka

DECEMBER 2025

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 18315

Separating Preferences from Endogenous Effort and Cognitive Noise in Observed Decisions

Christian Belzil
CREST, CNRS, Paris Polytechnic Institute, IZA and CIRANO

Tomáš Jagelka
University of Bonn, Dartmouth College, CREST-Ensae and IZA

DECEMBER 2025

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

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

We develop a micro-founded framework to account for individuals' effort and cognitive noise which confound estimates of preferences based on observed behavior. Using a large-scale experimental dataset we find that observed decision noise responds to the costs and benefits of exerting effort on individual choice tasks as predicted by our model. 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.

JEL Classification: D91, C40

Keywords: economic preferences, latent attributes, stochastic choice models, endogenous effort, cognitive noise, complexity, experimental design, achievement tests

Corresponding author:
Tomáš Jagelka
Universität Bonn
Regina-Pacis-Weg 3
53113 Bonn
Germany
E-mail: tjagelka@uni-bonn.de

* 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, Charlie Holt, Nicolas Jacquemet, Erzo F.P. Luttmer, Sebastian Olschewski, Jawwad Noor, Antonio Penta, Philipp Schirmer, Nikolaus Schweizer, Douglas Staiger, Robert Sugden, John Rust, Erik Wengström, and Florian Zimmermann for helpful discussions, comments, and suggestions. We also thank participants at seminars at the University of Bonn, CERGE-EI, Duke University, the Economic University of Bratislava, Laval University, MIT, Universitat Pompeu Fabra, Université Sorbonne at Paris 1 and Paris School of Economics, Tilburg University, the University of Virginia, 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 SPUDM 2025 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.

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 and cognition. 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, decision noise remains a potential source of bias, and of apparent preference instability, when the analyst naively equates an observed choice with the decision-maker's preference.[1] Our main contribution is to develop and estimate a micro-founded stochastic choice model which separates the signal on preferences in observed choices from inattention due to endogenous effort decisions and exogenous cognitive noise. 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 preferences of their members in a quantitative way (see, e.g., Goossens et al., 2023). We model task complexity as one of the inputs into an individual's endogenous effort decision, which in turn impacts the probability that the individual will make a choice in line with their true preference. Our model implies a general relationship between bias in preference estimates (risk, time, social, etc.) and factors which reduce the benefits of exerting effort, or increase its costs, when decision errors due to inattention are 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 formula for predicting bias in preference estimates from choice data and demonstrate its effectiveness even in preference elicitation tasks with many choice options.

Our analysis is in line with recent research in psychology and economics which recognizes that effort and imperfect perception of decision attributes influence observed measures even in controlled settings.[3] A key innovation is that we separate decision noise into two conceptually dis-

[1] Another extreme, equally dangerous, is to take the apparent preference instability as evidence that observed decisions are a pure manifestation of decision noise and true preferences either do not exist or play no role in them.

[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).

[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

tinct components: (i) endogenous (in)attention which responds to the costs and benefits of making a choice on a task according to an individual's true preference and can be reduced by applying more effort and (ii) exogenous cognitive noise which remains even at full effort, is outside of an individual's control within the timeframe of observation, and may only be reduced over a longer period of time as an individual improves their degree of self-knowledge regarding the preference of interest.[4] We model cognitive noise as a preference shock that influences the individual's assessment regarding which is his preferred (higher expected-utility) alternative[5], while exerted effort, relative to a task's individual-specific difficulty, modulates the probability with which the individual is able to select his preferred alternative on a given task.

In order to grasp the intuition behind our estimation strategy, it is useful to make an analogy with standard factor analysis methods used to infer an individual-specific attribute (often cognitive skill). These extract the latent factor 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 relative effort probabilities which act as choice specific weights.[6] These are then used to distinguish between choices deemed informative of structural parameters (risk aversion in our case) and choices that are largely random and thereby less informative of risk aversion. However, a major difference with statistical factor analysis is that the weights associated to each choice are endogenously determined by the individual. While we apply it to a setting devoted to measuring risk aversion, our approach is general and may be used in any context where the econometrician can access data on individual choices that are exerted with varying stakes and/or require different levels of cognitive load.

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

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]We model cognitive noise as the *epistemic uncertainty* of an individual with regards to his true preference. This is a complementary approach to assuming that cognitive noise manifests itself as an imperfect perception of task attributes (e.g., payoffs or probabilities) as in Woodford (2020), which is plausible in particular for attributes which occur infrequently (see Frydman and Jin, 2022). We view the imperfect perception of a latent preference as empirically more relevant than uncertainty regarding well-defined payoffs of a reasonable magnitude. However, preference shocks and payoff shocks can be shown to yield equivalent choice probabilities in an expected utility framework under reasonable assumptions. Note that our model is clearly distinct from a strand of recent work which emphasizes noise as the main driver of observed choices (e.g., Vieider, 2024) in that we assume the existence of standard economic preferences and utility functions. Yet another 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).

[6]Throughout the paper we focus on relative effort (i.e., effort relative to a task's difficulty) which, when normalized to the [0;1] interval, can be interpreted as the probability than an individual's choice is informative of his latent preference. This is in contrast to some of the existing literature, which is concerned with effort understood as the amount of cognitive resources (such as time spent thinking) exerted in an absolute sense (e.g., Gonçalves, 2024).

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

of such choice tasks in this experiment. While 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, they 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 relative effort increase with task complexity, with low relative stakes, and with fatigue—instances in which the costs of making a 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. 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 percentage falls to only 20% in choices on tasks of the more complex design. Accordingly, heterogeneity in risk preferences accounts for 90% of the explained cross-sectional variation in an individual's average choices on tasks of the simpler design but only for 50% on the more complex tasks (the other half is largely noise due to inattention).

Incorporating endogenous effort improves model fit by approximately 15%. Accounting for endogenous effort is particularly crucial when observed choices contain a lot of noise. While the distribution of the coefficients of risk aversion estimated using 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 when 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 quanti-

[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.

[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

fies 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. In addition, we find that women's effort decisions are less sensitive to choice stakes and fatigue but more sensitive to task complexity. This suggests a nuanced pattern of bias in preference estimates for different demographic groups which should be further explored.

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 estimates of risk aversion and relative effort obtained from the observed binary choices, our model correctly predicts not only the direction of bias due to insufficient effort at the individual level on the multiple choice tasks but also explains much of its cross-sectional variation.

Finally, we show that estimated propensity to exert sufficient effort 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). Transposing our results into the contemporaneous 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).

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

We contribute to the recent literature that links discrete choice models with concepts of Costly Reasoning (Alaoui and Penta, 2022), Rational Inattention (Steiner, Stewart, and Matějka, 2017;

lotteries with a 50% probability of receiving either payment.

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), Noisy Cognition (Vieider, 2024), Imperfect Self-Knowledge (Jagelka, 2024; Dohmen and Jagelka, 2024; Falk, Neuber, and Strack, 2021), or Limited Attention (Barseghyan, Molinari, and Thirkettle, 2021). As such, this paper enriches the broader domain of behavioral inattention summarized by Gabaix (2019).

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.[11]

Although the RPM is intrinsically monotone, 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.[12] 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

[11] Efforts to establish conditions under which aRUM applied to risk preference estimation can be monotone are ongoing (see, e.g., Keffert and Schweizer, 2024)

[12] 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.

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. 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 cost of 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.[13]

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.[14]

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). We call this residual exogenous decision randomness *cognitive noise*. 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,

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

[14] 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).

an individual's estimated degree of cognitive noise, unlike the propensity to pay sufficient attention, is independent of task design. This is what one would expect if the scale of a preference shock captures an individual characteristic such as imperfect self-knowledge.[15]

3 Model

Before providing technical details, let us exposit the general set-up of the model: An individual makes choices on binary tasks designed to elicit a preference. Each choice provides information about the individual's latent preference of interest if he takes the task seriously and has at least some self-knowledge regarding his preference.

When an individual is presented with a choice task, he takes in the readily and effortlessly available characteristics of the options among which he has to choose. He then decides how much effort to expend on making the choice. The amount of exerted effort *relative to* the task's individual-specific difficulty ($E^R_{i,l}$) will influence the likelihood that individual i will in fact choose the utility maximizing alternative on task l (henceforth the "*preferred option*"), given his relevant latent preference and an error shock representing cognitive noise.[16] As effort does not have a natural scale, we normalize $E^R_{i,l}$ to the [0;1] interval without loss of generality. At the top end of the scale, when $E^R_{i,l}$ approaches 1, the individual will choose his preferred option with close to certainty. At the bottom end of the scale, when $E^R_{i,l} = 0$, the individual's choices will reflect pure noise. At intermediate levels of relative effort, we assume that the probability of choosing the preferred option is monotonically increasing in exerted effort. $E^R_{i,l}$ can thus be interpreted as the probability that choice l is informative of individual i's latent preference.[17] It can be seen as the "informativeness" weight that the econometrician, intent on inferring the preference of interest, would want to assign to the particular choice.

Consider a task involving a choice between two options: Y and X. An observed choice of Y can come about in two ways: (i) the individual prefers Y and exerted sufficient effort on the task to select his preferred option, or (ii) the individual was not paying sufficient attention and chose option Y randomly. Our estimation strategy accounts for the fact that the observed choice reveals the individual's latent preference only in the first case.

We can write the probability that individual i *chooses* option Y on a binary choice task l as:

[15]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.

[16]The effort decision may be taken at an implicit level. Relative effort, $E^R_{i,l}$, could be further decomposed into effort exerted by individual i on task l, and the task's individual-specific difficulty. However, *relative effort* is the object relevant to the econometrician as it determines the probability with which the individual will choose the option that he truly prefers.

[17]$E^R_{i,l}$ can also be understood as the probability that individual i exerts *sufficient effort* on task l to be able to *reliably* choose the option that he prefers. The actual probability of choosing either option can never fall below the probability with which the option is chosen if the individual is choosing randomly between the options.

$$p(YC_{i,l} = 1) = E^R_{i,l} \cdot p(YP_{i,l} = 1) + (1 - E^R_{i,l}) \cdot r_{Y,i} \tag{1}$$

where $p(YC_{i,l} = 1)$ is the probability that individual i *chooses* option Y on task l; $p(YP_{i,l} = 1)$ is the probability that individual i *prefers* option Y on task l; $E^R_{i,l}$ is exerted effort of individual i relative to task l's subjective difficulty normalized to the [0;1] interval; and $r_{Y,i}$ is the individual's "effortless" randomization strategy which determines the probability with which individual i picks option Y when exerted effort is insufficient to reliably make a choice in line with his true preference. A reasonable default value is $r_{Y,i} = 0.5$, i. e., an individual who exerts minimal relative effort 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

In the previous section we established that the econometrically interesting object is the amount of effort that an individual will choose to exert *relative to a task's difficulty* because it is closely related to the likelihood that the individual will be able to reliably choose the option which he prefers. The individual will want to increase relative effort ($E^R_{i,l}$) when the benefits of choosing the preferred option are high and the costs are low.[18]

The *benefits* of choosing the preferred option will generally be increasing in the rewards associated with a choice and in the difference in the attractiveness of the choice options. The *costs* of choosing the preferred option will generally be increasing in the amount of effort required and in the cost to the individual of exerting that amount of effort. Denote B_l the vector of readily and effortlessly available characteristics which pertain to the perceived *benefits* of exerting sufficient effort such that an individual is able to reliably choose his preferred alternative. 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. Let us assume that individuals act according to *net* perceived benefits.

As mentioned in the previous section, because effort has no natural scale, we can normalize $E^R_{i,l}$ to the [0;1] interval without loss of generality. The effort that individual i chooses to exert when faced with task l, relative to that task's difficulty, can thus be expressed as follows:

$$E^R_{i,l} = \Phi(b_{0,i} + b_{1,i} \cdot B_l - b_{2,i} \cdot C_l) \tag{2}$$

where Φ is the cumulative distribution function for a standard normal distribution, $b_{0,i}$ denotes the intercept which captures individual differences in baseline propensity to exert relative effort in the analyzed choice tasks (e.g., due to differences in personality or variability in how difficult

[18]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.

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.[19]

The normalized $E_{i,l}^R$ can be interpreted as the probability that individual i exerts sufficient effort on task l to be able to *reliably* choose the option he prefers on that task, i.e, it is the probability that the choice will be informative of the individual's latent preference to the econometrician.[20]

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). 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).[21]

The residual decision 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.[22] The discounted expected utility that an individual i derives from a choice option thus depends on choice characteristics,

[19] While in principle $b_{1,i}$ and $b_{2,i}$ are unbounded, we limit them to the [-5;5] interval in estimation in order to avoid numerical issues which arise as one approaches the limits of the standard normal cumulative distribution function.

[20] Indeed, Equation 2 can also be derived from an alternative set of assumptions: The individual is simply deciding whether or not to take the decision task seriously, i.e., whether or not no exert sufficient effort, relative to the task's individual specific difficulty, such that he is able to reliably choose according to his latent preference. When the effort decision itself is noisy, it yields a probability of exerting sufficient effort.

[21] A person who is unsure of their true preference may randomize within an interval of uncertainty which depends on individual characteristics (Jagelka, 2024).

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

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 over option X when $DEU_i(m_y,n_y,o_y;\Psi_i;\varepsilon_i) > DEU_i(m_x,n_x,o_x;\Psi_i;\varepsilon_i)$.[23] 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 the probability with which an individual is able to convert the preference into an actual choice.

3.c Application to Risk Preference Elicitation

The general model is easily adapted to choice under risk:

Let us consider an example in which the researcher observes individuals making binary choices between lotteries, as is the case in our dataset described in Section 4. Shifters in the benefits of exerting effort should target the magnitude of the lottery payoffs, or the relative attractiveness of the two lotteries. Shifters in the costs of exerting sufficient effort should alter the amount of effort needed to reliably select an individual's preferred lottery, or the disutility associated with a "unit" of exerted effort.[24]

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., $r_{Y,i} = 0.5$.

Let $U_i(a)$ represent the utility which an individual obtains from a dollars. Define the coefficient

[23] 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 discounted expected utility determine the preferred choice.

[24] Given effort shifters available in our data, we let (i) the effort cost vector C_l consist of a dummy for task design (allowing for tasks of the more complex design to have a different baseline effort requirement than tasks of the simpler design) and an indicator for task order (allowing for the cost of a *given* amount of effort to change with fatigue and/or learning); (ii) the effort benefit vector B_l consist of the percentage difference in the expected payoffs offered by each lottery (reflecting the stakes associated with not making a mistake).

of relative risk aversion $\theta_i = \frac{-a \cdot U''(a)}{U'(a)}$.[25] 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.[26] 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.[27] The individual uses the shocked (or instantaneous) value of risk preference $\theta_i + \varepsilon_i$ to 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

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

[25] We restrict θ_i to the (wide) range of the coefficient of relative risk aversion covered by the available elicitation tasks described in Section 4, so $\theta_i \in (-2, +5)$, see Section B.b of the Online Appendix..

[26] As explained by Loomes and Sugden (1995): "the stochastic element derives from the inherent variability or imprecision of the individual's preferences, whereby the individual does not always know exactly what he or she prefers. Alternatively, it might be thought of as reflecting the individually small and collectively unsystematic impact on preferences of many unobserved factors." 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).

[27] For closed form solutions of the choice probabilities under the alternative random utility specification with additive shocks (aRUM), please see Section B.a of the Online Appendix.

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 $\overline{\varepsilon}_{i,l}$, the maximum value which still satisfies the inequality expressed in Equation (9):

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

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:

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

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 variable with mean θ_i and standard deviation σ_i, both of which are parameters to be estimated.[28] 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.

Combining Equations 1, 2, and 12 while assuming than when an individual is choosing randomly he randomizes between the available options with equal probability, we obtain the closed-form expression for the probability that individual i *chooses* the riskier option Y on choice task l:

$$p(YC_{i,l} = 1) = \Phi(b_{0,i} + b_{1,i} \cdot B_l - b_{2,i} \cdot C_l) \cdot \Phi\left(\frac{\theta_l^{eq} - \theta_i}{\sigma_i}\right) + [1 - \Phi(b_{0,i} + b_{1,i} \cdot B_l - b_{2,i} \cdot C_l)] \cdot 0.5 \tag{14}$$

Both σ_i and $E_{i,l}^R$ impact the consistency of an individual's repeated observed choices. However,

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

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 his instantaneous preference can vary somewhat from question to question, in the absence of decision mistakes an individual would be choosing the preferred expected utility maximizing option given his current (shocked) risk preference. On the other hand, by electing not to exert full effort 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 Decision Noise Parameters

Both σ_i and $E^R_{i,l}$ 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 probability of exerting sufficient effort on all experimental tasks, $E^R_{i,l}$.[29] 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 that time. Thus the individual would choose the option which gives him lower expected utility $\frac{1-E^R_{i,l}}{2}\%$ 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 relative effort could explain such choices in our model and $E^R_{i,l}$ would therefore trivially be identified from such choices.

The constant relative 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 as the residual noise after stripping away the uniform noise component due to insufficient effort provision.

More generally, $E^R_{i,l}$ 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 tasks with a θ^{eq}_l *close to*, or *far away from*, an individual's latent risk

[29] Recall that the normalized $E^R_{i,l}$ can be interpreted as the probability that individual i exerts sufficient effort on task l to be able to reliably choose the option he prefers on that task.

preference θ_i.[30] 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 reliably 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.[31]

3.e Estimation

Individual i's contribution to the likelihood based on his choice on lottery choice task l is:

$$p(YC_{i,l} = yc_{i,l}) = p(YC_{i,l} = 1)^{YC_{i,l}} \cdot p(YC_{i,l} = 0)^{1-YC_{i,l}} \tag{15}$$

where the probability $p(YC_{i,l} = 1)$ that individual i chooses option Y on task l is given by Equation 14.

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

$$L_i = \prod_{l=1}^{L} p(YC_{i,l} = yc_{i,l}) \tag{16}$$

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

[30] 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.

[31] 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.[32] 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 other than 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.[33] 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 of choices per individual 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. Characteristics of the lotteries that are readily and effortlessly available to each individual, and therefore factor into the effort decision, are: task complexity and order (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 their 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, each containing 10 choice

[32]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).

[33]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.b.

tasks. 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). 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 an 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 in each ordered group is also simple as the higher payment is received with certainty and thus there is a dominated option.[34] 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, each containing 5 choice tasks. Once again, in each group of tasks, 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 on latent risk preference and noise due to endogenous 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.[35] 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.[36] In a deterministic world, the sMPL tasks should allow

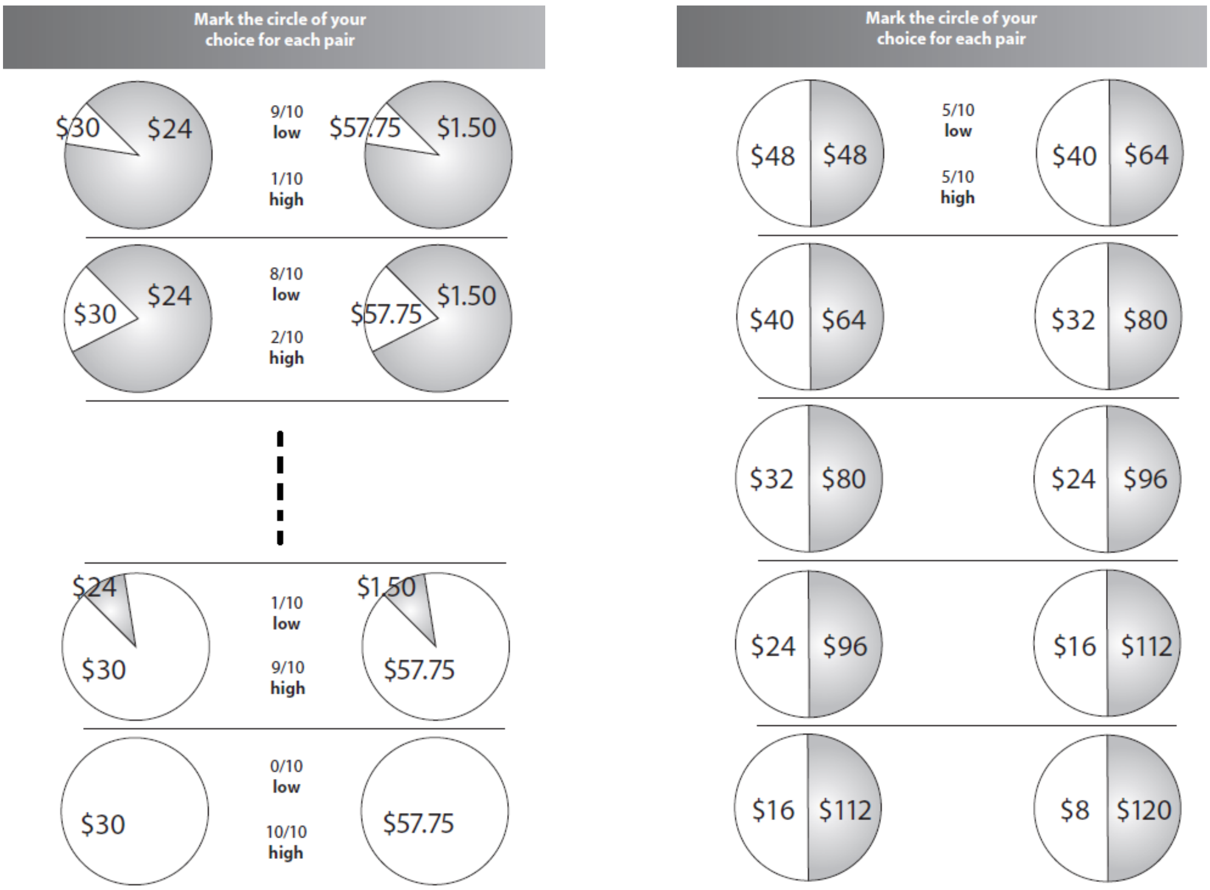
[34] In the last row of all three sets of sMPL questions designed to measure risk aversion, both lotteries offer the higher payment with certainty. Because no value of risk aversion can justify a preference for lottery X, it is dominated by lottery Y.

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

[36] 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



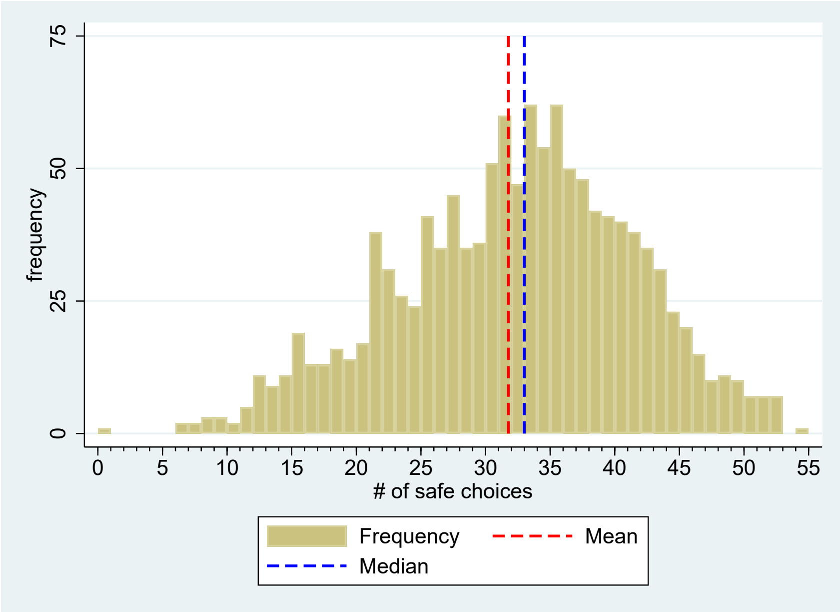
for the identification of an interval for an individual's risk aversion while the cMPL tasks should permit the refinement of this interval. Furthermore, while the sMPL tasks focus on the most common range of risk preferences (up to a coefficient of relative risk aversion of 1.37 under CRRA utility), 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 rigorous 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.[37] 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 insufficient attention.[38] 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.c.iii.

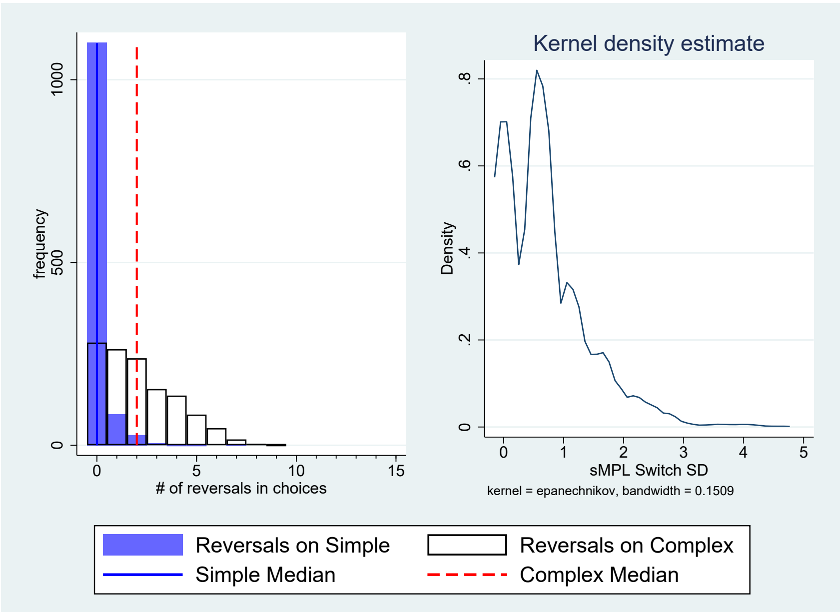
Figure 3 plots the distributions of reversals *within* a given MPL and of inconsistency in switching points *between* comparable MPLs. It reveals that while some reversals are observed on sMPL tasks, most of the action takes place on cMPL tasks. While almost 90% of individuals exhibit no reversal behavior 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 to

[37] 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.

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

latent risk preference on a task 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



Notes: The left panel shows the histogram of the number of choice reversals, within an order list of lottery choice tasks of the simple sMPL design (blue bars) and of the more complex cMPL design (transparent bars), exhibited by individuals in our dataset. The right panel plots the distribution of the standard deviation of an individual's switching points across MPLs on which the switching point is predicted to be the same for a given coefficient of relative risk aversion under CRRA utility. The distribution is smoothed through kernel density estimation.

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 of the Appendix.

5 Empirical Results

5.a Representative Agent Model

We first estimate a representative agent model to obtain a baseline comparison for our individual-by-individual estimates.

Our model with endogenous effort results in an approximately 15% improvement in the log-likelihood relative to the model with no effort.[39] The model with endogenous effort dominates also using the Akaike and Schwartz information criteria which include a penalization for the number of parameters.

Furthermore, omitting endogenous effort results in an estimated coefficient of relative risk aversion which is more than 50% higher than when endogenous effort is taken into account (1.1 vs. 0.67). Apparent preference instability (cognitive noise) is also magnified in that case. Finally, we note that allowing endogenous effort decisions to differ by sex appears fruitful: the interaction effects are statistically significant and model fit further improves. These results are summarized in Appendix Table A.2. We explore these basic findings in more depth in the rest of the paper, where we allow all structural coefficients to be individual specific.

5.b Individual Estimates

Individual-specific 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 relative effort required for observed choices on 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 relative effort for being able to *reliably* choose according to underlying preferences, averaged over the 55 tasks that each individual faced. Figure 4 plots the parameter distributions.[40]

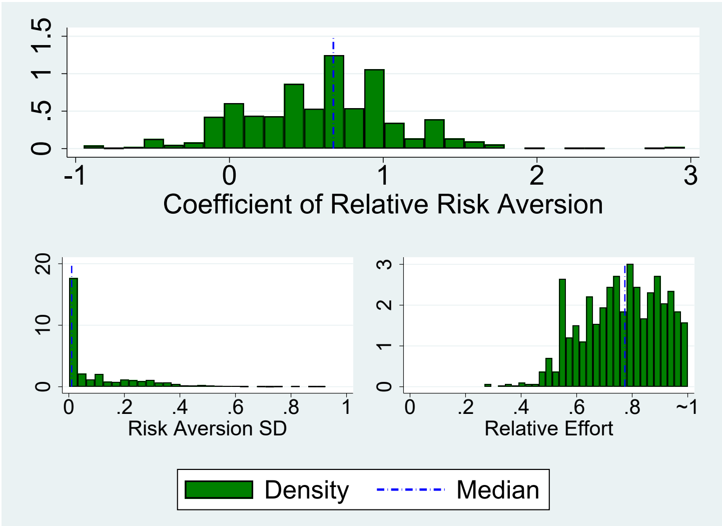
In order to put these results in context, it is helpful to compare them to existing estimates.[41] The obtained values of the coefficient of relative risk aversion are broadly in line with the previous literature (see e.g., Holt and Laury, 2002; Andersen et al., 2008; Apesteguia and Ballester,

[39] For the purposes of this exercise, we exclude the 3 choice tasks which feature a dominated choice from the analysis as a simple random preference model cannot explain such choices.

[40] 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.

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

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



2018; Apesteguia, Ballester, and Gutierrez, 2020; Jagelka, 2024).[42] 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 individual.[43] 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.c.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.d 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.c Endogenous Effort

Following our theoretical model, we allow the effort parameter to depend on readily and effortlessly available task characteristics which shift the costs and benefits of choosing the preferred option. 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).

[42]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.

[43]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”.

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, $E^R_{i,l}$, for the median individual.[44] The marginal effect of increasing relative stakes by one standard deviation averaged across all 55 lottery choice tasks is a 7% increase in $E^R_{i,l}$ whereas increasing fatigue by one standard deviation results in a 2% decrease in $E^R_{i,l}$.[45] These results are congruent with our estimates from the representative agent model. Appendix Table A.3 shows that women on average exert higher relative effort on the analyzed experimental tasks (i.e., their choices are on average more informative of latent preferences). As a group, women are less sensitive to choice stakes and fatigue, but more sensitive to task complexity.

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 revealed 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 relative effort decision and explain its determinants.

5.c.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 decision noise parameters σ_i and E^R_i.[46] 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 across individuals. 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

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

[45]Task order could also be associated with learning, which would presumably work to *decrease* the cost of effort required to reliably make a choice in line with one's latent risk aversion. We interpret the estimated negative effect of task order on relative effort as suggesting that the effect of fatigue outweighs any potential benefits from learning. We calculate all 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 $E^R_{i,l}$ given each lottery's actual characteristics and the counterfactual $E^R_{i,l}$ if the design were flipped to cMPL, or if relative stakes or fatigue were increased by one standard deviation.

[46]We obtain an individual's propensity to exert sufficient effort E^R_i as an average of the estimated task-specific relative effort propensities $E^R_{i,l}$. We can average $E^R_{i,l}$ alternatively over tasks of the simpler design to obtain $E^R_{i,sMPL}$ or over tasks of the more complex design to obtain $E^R_{i,cMPL}$.

parameters into parts explained by the preference parameter and by the decision noise parameters. This lets us compare their relative explanatory power, expressed as a percentage. Decision noise parameters are further broken down into exogenous cognitive noise and endogenous relative effort. This allows us to provide empirical evidence on the separate identification of the two types of decision noise 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 variation in latent risk preference across individuals 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 the more complex tasks.[47] This is yet another indicator that choices on the cognitively less demanding task design are driven by individuals' risk preferences to a much greater extent then choices on the cognitively more demanding task design.

Table 1: Variation in Average Behavior on Lottery Choice Tasks Attributed to Preference vs. Decision Noise 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%
Decision Noise Parameters		10.9%	11.8%	46.4%	100.0%	99.7%	100.0%	99.5%
- Cog. Noise		0.4%	0.8%	0.1%	0.0%	7.2%	0.1%	59.7%
- Relative 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. Socioeconomic variables include parents' level of education and income. Due to the presence of missing values, this regression has 888 observations. 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 ordered group of lottery tasks 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 1,135 observations or over 90% of the sample.

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 relative effort on both task designs. This is consistent with the finding that the median individual exhibits stable risk preferences. Choice inconsistency on lottery tasks is thus largely due to

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

mistakes due to endogenous effort decisions. However, cognitive noise, captured by the estimated 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. One can see that while cognitive noise and propensity to exert sufficient effort both explain randomness in observed decisions, they manifest through distinct patterns of choice inconsistency 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.[48]

5.c.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 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 θ_l^{eq} 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* θ_l^{eq} *for a given choice task* (henceforth referred to as the "threshold dummy") is the main predictor of an observed choice on that task.[49] 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

[48]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.

[49]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.

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
Relative Effort	R2	0.00	0.00	0.01	0.24	0.16	0.18
Relative 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. Due to the presence of missing values, regressions with demographic and socioeconomic variables include 52,360 observations (out of a 67,320 observations total) when all choices are considered. The "Threshold Dummy" is equal to one if the estimated coefficient of relative risk aversion is below the indifference threshold for a given task. "Relative Effort" is a task specific probability that an individual will exert sufficient relative effort to be able to choose his preferred option (including the cognitive noise shock) given task characteristics and his estimated relative effort 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 endogenous relative effort (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.[50] The threshold dummy and its interactions with the remaining structural parameters contribute minimally. This provides empirical support for the assumption that when an individual exerts low relative effort, he will randomize between the safe and the risky lottery with equal probability.[51] Finally, it is noteworthy that the 18 included demographic and socioeconomic variables together predict neither observed nor wrong choices.

5.c.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.

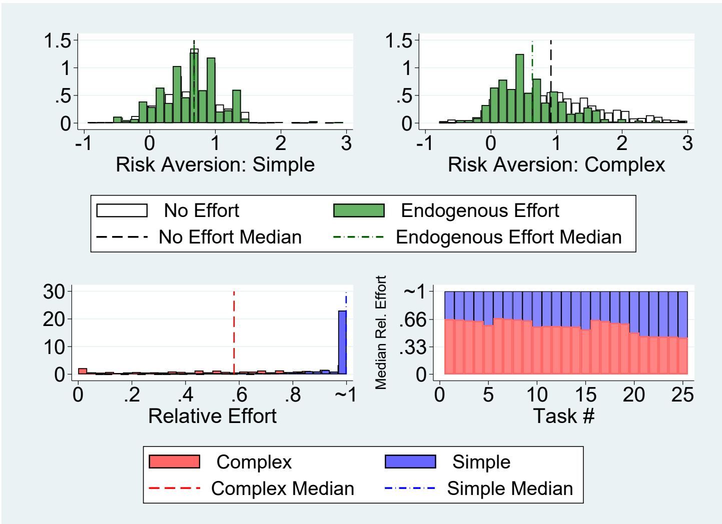
In the context of our experiment, relative choice stakes and fatigue meaningfully influence effort decisions only on the more complex tasks (i.e., their estimated average marginal effect for the median individual on the simpler tasks is zero). This is easily discernible from the bottom right

[50] "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.

[51] In other words, knowing whether or not an individual actual prefers the safe or the risky lottery on a given task does not help us predict whether he is likely to make a mistake on it.

histogram of Figure 5, which plots estimated relative effort for the median individual on the first 25 tasks of each design. Furthermore, when we average estimated relative effort 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 relative effort such that they are able to choose their preferred option almost all of the time on the simpler tasks. In contrast, the median individual's exerted relative effort enables him to reliably select the preferred option only 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.[52] 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).[53]

Figure 5: Distributions of Structural Parameters by Task Design



Notes: The top two panels plot distributions of the estimated coefficient of relative risk aversion in the sample, alternatively using the model with endogenous relative effort (green bars) and without it (transparent bars). The top left panel plots estimates based on choices on tasks of the simple design while the top right panel plots estimates based on choices on tasks of the more complex design. The bottom left panel plots the distribution of estimated relative effort exerted by individuals in the sample, averaged alternatively over tasks of the simple design (blue) or of the more complex design (red). The bottom right panel shows the relative effort exerted by the median individual on tasks 1-25 of the simple design (blue) and on tasks 1-25 of the more complex design (red).

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 alternative more often than the safer one.[54] We test this hypothesis formally. For each individual, we

[52]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.d.

[53]The estimated coefficient of relative risk aversion using the more complex choice tasks is 0.6 when endogenous relative 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.

[54]When actual risk preference leads an individual to choose relatively many riskier options, random errors are

first calculate the difference between the estimated coefficient of relative risk aversion obtained from the noisy complex design when the effort parameter is omitted and when it is included. This is the (upawards) bias in estimated 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 our estimate of his true 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 Individuals' Predicted Percentage of Risky Choices and Relative Effort on cMPL Tasks

	Estimated Bias in the CRRA Coefficient of Relative Risk Aversion	
Variables	(1)	(2)
Predicted % Riskier	2.70***	0.33***
	(0.82)	(0.093)
Relative Inattention		-0.14
		(0.09)
Predicted % Riskier * Relative Inattention		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 to determine this bias. "Predicted % Riskier" 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 $\hat{\theta}_i$. "Relative Inattention" is the converse of average relative effort exerted on the more complex tasks, i.e. $1 - E^R_{i,cMPL}$ in the notation of our model. It represent the 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 relative 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 relative 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 sufficient effort on the more complex tasks 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 0.77 increase in the (upwards) bias of the estimated coefficient of relative risk aversion.[55] It can be understood as the effect of design imbalance at the individual level.

Given that task complexity is the key determinant of endogenous relative effort in our setting, our

more likely to flip a truly preferred choice of a risky option to safe than the converse. This implies fewer observed risky choices than justified based on the person's latent risk preference and overestimation of risk aversion if decision error is not properly taken into account.

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

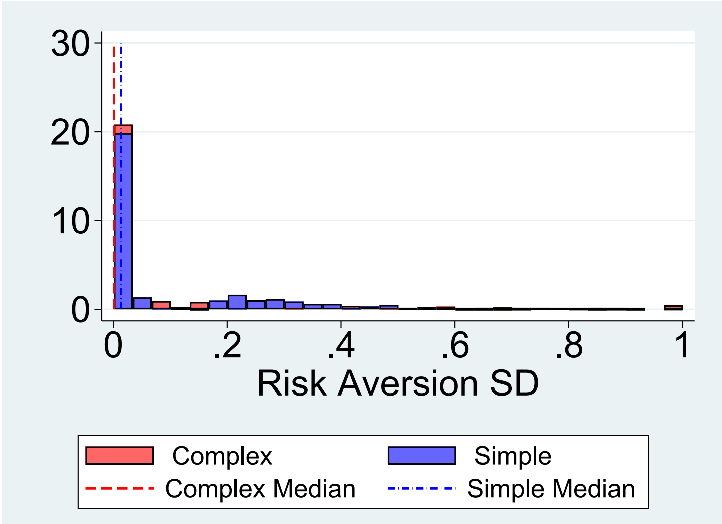
findings predict a general relationship between elicitation task complexity and bias in preference estimates. As an illustration, consider a hypothetical set of multiple ordered lists of tasks, each consisting of repeated binary choices eliciting the same parameter of interest (e.g., preference for risk, 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 least to the most complex list. As exerted relative effort 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 he truly is (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 the individual is less risk averse then he truly is (a downward bias).

5.d 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 (see, e.g., Jagelka, 2024). 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. 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. Apparent preference instability (cognitive noise) estimated using only tasks of the same design is low. Furthermore, the distribution of cognitive noise estimated separately on the simpler and more complex tasks is similar, in contrast to the distribution of estimated relative effort (see Figure 6 and the bottom left histogram of Figure 5, respectively). The fact that cognitive noise is similar across task designs while mistakes due to inattention vary, suggests that the stability of preferences is an individual characteristic (and can reasonably be considered as exogenous within the timeframe of an experiment) while decision errors are due to endogenous effort decisions, responsive to incentives.

Figure 6: Distribution of the Estimated Cognitive Noise Parameter by Task Design



Once the decision to exert effort is incorporated into the model, the median individual has stable estimated risk preference 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 on which individuals exert low relative effort.[56] This suggests that the failure to account for differences in situations may in general result in an overestimation of preference instability or cognitive noise.

The inclusion of a properly parametrized relative effort parameter seems recommendable if a researcher uses information on choices arising in different situations, which are likely to entail varying degrees of decision noise. We show that modeling inattention as a function of a few readily available attributes is able to account for differences in situations in our context 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. Nevertheless, 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.

[56] In our dataset, these are largely tasks of the more complex design.

6 External Validity and Out-of-Sample Predictive Power

6.a External Validity

While the internal validity of our model is well documented, an intriguing questions remains: Does the estimated individual propensity to exert sufficient effort in a relatively low-stakes experimental setting capture an individual's broader tendency to exert effort and predict external outcomes?

To answer this question, we make use of the pre-experiment survey which contains two different measures of student achievement: the International Adult Literacy Survey (IALS) 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.[57] 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 (see, 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 confounders, 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*. In contrast, individual grades 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*.

Table 4 shows that effort estimated from the relatively low-stakes lottery choice tasks we study predicts both numeracy scores and high school GPA, even after controlling for self-reported skills, personality, and sex.[58] 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, the estimated marginal effect is meaningful in magnitude. Increasing effort by one standard deviation, holdings self-reported skills, personality, and

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

[58] To test the predictive power of effort at an individual level, we use our estimate of an individual's relative effort averaged across all 55 binary lottery choice tasks faced by the individual.

sex constant, is predicted to increase an individual's numeracy score by 0.12 standard deviations and their high school GPA by 0.09 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.[59] 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 (see, OECD, 2019). A 0.12 standard deviation increase therefore corresponds to an increase of 12 points on the PISA test. 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, see OECD, 2010), 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]

Table 4: Predictive Power of Estimated Relative Effort on the IALS Achievement Test and High School GPA

VARIABLES	(1) IALS	(2) HS GPA
$\hat{E}_i^R$	0.12***	0.09***
	(0.03)	(0.02)
Cognitive Skills	x	x
Non-Cognitive Skills	x	x
Risk Preference	x	x
Sex	x	x
Constant	0.05	-0.15***
	(0.04)	(0.04)
Observations	1,224	1,224
R-squared	0.19	0.29

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. $\hat{E}_i^R$ is our estimate of an individual's relative effort averaged across all 55 binary lottery choice tasks faced by the individual. 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 relative effort model based on all 55 binary lottery choice tasks.

Online Appendix Table B.4 provides additional interesting insights on the skills and preferences which impact numeracy achievement tests and high school GPA. For example, it shows that, as expected, self-reported math skills are the single most important predictor of numeracy scores. It also shows that conscientiousness is the single most important predictor of high school GPA. Finally, we can see that our estimate of effort is a stronger predictor of the low stakes achievement test scores than of the high stakes high school GPA, both in terms of its estimated marginal effect

[59] 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).

[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.

and the share of explained variation in the outcomes. This raises the intriguing possibility that a tendency to exert effort in low-stakes and high stakes environments are separate individual attributes which merit further study.

6.b Out of Sample Predictive Power: 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 lottery task (MCLT) combines the lotteries from an ordered group of 5 binary choice tasks of the more complex "cMPL" design. See Figure 7 for an example combining the 5 cMPL tasks from Figure 1 into a single task.[61]

We use the MCLT tasks to test the predictions of our model in a related but different setting. To link our results with the non-structural literature on preference estimation, in this subsection we do no further estimation. We simply take our estimates of individual effort propensities and risk aversion from the binary choice tasks, along with our structural model with endogenous effort, to this additional multiple choice data. In particular, we wish to evaluate whether: (i) our estimates of an individual's risk aversion from the binary choice tasks predict risk aversion implied by his choices on the multiple choice tasks; (ii) our estimates of an individual's exerted relative effort from the binary choice tasks predict the consistency of his choices on the multiple choice tasks; (iii) our combined estimates of an individual's risk aversion and exerted relative effort from the binary choice tasks predict bias in the coefficient of relative risk aversion implied by the individual's choices on the multiple choice tasks when decision noise is not taken into account, relative to our estimate of his true risk aversion.

We calculate the coefficient of relative risk aversion implied by an individual's choices on the MCLT tasks using the theoretical indifference thresholds between adjacent lotteries (see Figure 7 for an illustration).[62] We call this coefficient of relative risk aversion *naive* because it because it does not take into account bias due to decision noise.[63] We denote it $\tilde{\theta}^N_{i,m}$ where the "squiggle" reflects the fact that the coefficient is calculated (rather than estimated) based on individual i's observed choice on MCLT task m. We also calculate the implied coefficient of relative risk aversion $\tilde{\theta}^N_{i,M}$ that would be *inferred jointly from all 5 individual i's observed multiple choice decisions* as a simple average of the $\tilde{\theta}^N_{i,m}$ implied by each of his 5 observed choices individually.

[61] The 5 resulting MCLT tasks are analogous to the design used by Eckel and Grossman (2002). The writing in blue was added for purposes of illustration of the indifference threshold method used below and was not shown to experiment participants.

[62] For example, if an individual were to choose the lottery which pays \$40 half the time and \$64 half the time from among the 5 lotteries in Figure 7, the naive estimate of his coefficient of relative risk aversion $\tilde{\theta}^N_{i,m}$ based on that choice would be $(2.97 + 1)/2 = 1.99$.

[63] A similar approach has been employed in previous studies when authors wanted to obtain a quantitative measure for risk aversion grounded in theory without estimating a structural model, (see e.g., Eckel and Grossman, 2008; Dohmen et al., 2010).

Figure 7: Lottery Choice Tasks - Multiple Choice cMPL design

We next obtain the bias in the coefficient of relative risk aversion calculated without taking endogenous relative effort decisions into account. It is the difference between naive $\tilde{\theta}^N_{i,m}$ and our estimate of the individual's true coefficient of relative risk aversion $\hat{\theta}_i$ from the binary choice tasks which takes decision noise into account. Finally, we obtain the expected bias in the naive $\tilde{\theta}^N_{i,m}$ *predicted* by our model to arise when endogenous relative effort is neglected. We do so by adapting Equation 3 to a setting with multiple choices.[64] By subtracting bias predicted by our model from the naive $\tilde{\theta}^N_{i,m}$, we obtain a de-biased coefficient of relative risk aversion $\tilde{\theta}_{i,m}$ for individual i's choice on multiple choice task m. We refer the interested reader to Section A.b of the Appendix for details of these calculations, including a simple formula to predict bias at the individual level (see Equation 18).

[64]Intuitively, the naive $\tilde{\theta}^N_{i,m}$ is a weighted average between the coefficient of relative risk aversion which would be inferred if the individual chose his truly preferred option (with sufficient exerted relative effort) and the average of the coefficients of relative risk aversion implied by all the available choice options (between which he would randomize with equal probability in case of low relative effort). Predicted bias is high when exerted relative effort is low and when the non-preferred available lottery options imply values of risk aversion far from the individual's true risk preference. Given our estimate that the median individual does not exert sufficient effort to make a choice in line with his latent risk preference approximately a third of the time on the more complex binary choice tasks–and the resulting bias in risk aversion estimates we document in Section 5.c.iii when endogenous effort decisions are not taken into account–we expect the level of risk aversion implied by individuals' decisions on the *even more complex* multiple choice data to be *even more biased*.

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 specific hypotheses: (i) our estimate of an individual's latent coefficient of relative risk aversion $\hat{\theta}_i$ based on binary choice lottery tasks will predict the coefficient of relative risk aversion implied by his choices on the multiple choice lottery tasks, *and* it will better predict the de-biased coefficient of relative risk aversion $\tilde{\theta}_{i,m}$ than the naive $\tilde{\theta}^N_{i,m}$ implied by the individual's choice on multiple choice task m when endogenous effort is not taken into account; (ii) individuals for whom we estimated lower exerted relative effort on the more complex binary lottery tasks will make more inconsistent choices also on the multiple choice lottery tasks; and (iii) bias predicted by our model will predict actual bias in the implied naive coefficient of relative risk aversion at the individual level *and* it will be a lower bound on actual bias because the multiple choice tasks are even more complex than the binary task of the cMPL design used for the relative effort estimates.

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 true coefficient of relative risk aversion after taking endogenous effort into account. The results are shown in Table 5 below. Perfect predictive power of the estimated parameters on individuals' decisions on the multiple-choice lottery tasks would imply a constant equal to 0 and an OLS coefficient of 1 on $\hat{\theta}_i$.[65]

Table 5: **Predictive Power of an Individual's Risk Aversion Estimate from Binary Choices**: Explaining the Naive and De-Biased Coefficient of Relative Risk Aversion Implied by Choices on each MCLT Task

	MCLT Decision 1		MCLT Decision 2		MCLT Decision 3		MCLT Decision 4		MCLT Decision 5	
	Naive $\tilde{\theta}^N_{i,m}$	De-Biased $\tilde{\theta}_{i,m}$	Naive $\tilde{\theta}^N_{i,m}$	De-Biased $\tilde{\theta}_{i,m}$	Naive $\tilde{\theta}^N_{i,m}$	De-Biased $\tilde{\theta}_{i,m}$	Naive $\tilde{\theta}^N_{i,m}$	De-Biased $\tilde{\theta}_{i,m}$	Naive $\tilde{\theta}^N_{i,m}$	De-Biased $\tilde{\theta}_{i,m}$
End. Relative Effort 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 multiple choice lottery task on our estimate of that individual's latent risk aversion $\hat{\theta}_i$. The "End. Relative Effort Model $\hat{\theta}_i$" is obtained by estimating our model with endogenous relative effort using individual i's choices on all 55 binary choice tasks. The naive $\tilde{\theta}^N_{i,m}$ 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 $\tilde{\theta}_{i,m}$ is obtained by applying the bias correction implied by Equation 18 to the naive $\tilde{\theta}^N_{i,m}$.

Our results confirm that $\hat{\theta}_i$ estimated on binary choice lottery 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 our endogenous relative effort model estimates to the coefficient of relative risk aversion implied by choices on the MCLT tasks. This makes sense as the out of sample decisions involve a different context and the indifference threshold calculation of $\tilde{\theta}_{i,m}$ is coarse. In line with our hypothesis, $\hat{\theta}_i$, which already accounts for potential bias in risk aversion

[65] Conversely, if the full-model coefficients were not predictive at all, the OLS coefficient on $\hat{\theta}_i$ should be zero and the constant would capture the average population coefficient of relative risk aversion inferred from the multiple-choice questions.

estimates due to insufficient effort, better predicts the coefficient of relative risk aversion implied by choices on the MCLT tasks once we apply our bias correction.[66] This holds for each of the 5 MCLT tasks in our dataset taken individually, and also when we consider an individual's choices on them jointly (see Table A.4 of the Appendix). Indeed, the average share of variation explained by $\hat{\theta}_i$ roughly doubles once we apply our bias correction. At the same time, the estimated OLS coefficient increases and becomes closer to 1, and the estimated constant falls and becomes closer to 0.

To test our second hypothesis we verify that individuals who exhibit higher decision noise on the binary choice tasks also exhibit higher decision noise on the multiple choice tasks. Specifically, we check whether those who have a lower estimated exerted relative effort on the more complex binary choice tasks also have a higher dispersion in the naive coefficient of relative risk aversion implied by their observed choices on the five multiple choice tasks. This is the case as the correlation between our estimate of the individual's exerted relative effort on the cMPL tasks and the standard deviation of the coefficient of relative risk aversion implied by individual i's choices on the five individual MCLT tasks is -0.21, statistically significant at the 1% level.

To test our third hypothesis, we estimate how well our model predicts actual bias in the naive coefficient of relative risk aversion that would be inferred from the MCLT tasks without taking endogenous effort decisions into account. We do so by regressing actual bias on the bias predicted by our model given an individual's estimated risk aversion and relative effort from the more complex binary tasks. While perfect predictive power would still imply a constant equal to 0, this time we would expect an OLS slope coefficient >1 on predicted bias. This is because we hypothesize that individuals should exert even lower relative effort on the MCLT tasks than on the binary tasks of the cMPL design and thus the predicted bias should be a lower bound on actual bias.[67]

Table 6 reveals that our model is indeed able to predict the bias in the naive $\tilde{\theta}^N_{i,m}$, which would be inferred from a person's choice on each individual MCLT task (see columns 1-5), as well as the bias in the average $\tilde{\theta}^N_{i,M}$ 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 MCLT choices jointly. Furthermore, we cannot reject the hypothesis that the slope coefficient is greater than or equal to 1. Indeed, the point estimate is 1.6 when considering all 5 MCLT choices together.

Taken together, these results illustrate that our model with endogenous relative effort generalizes to a setting with multiple choice options. Our estimates of individuals' risk aversion and relative effort predict out of sample behavior on choices between multiple lotteries. A simple for-

[66] In other words, $\hat{\theta}_i$ better predicts the de-biased $\tilde{\theta}_{i,m}$ than the naive $\tilde{\theta}^N_{i,m}$.

[67] Recall that the MCLT tasks are even more complex than the binary cMPL tasks. This raises the cost of effort needed to choose according to an individual's true preference. It therefore lowers relative effort that the individual chooses to exert in our model, increasing the likelihood of decision mistakes.

Table 6: Actual vs. Predicted Bias in the Coefficient of Relative Risk Aversion Inferred from Choices on MCLT tasks without Taking Endogenous Effort Into Account

VARIABLES	MCLT Decision 1	MCLT Decision 2	MCLT Decision 3	MCLT Decision 4	MCLT Decision 5	**MCLT Average**
			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 for a given decision on a multiple choice lottery task (Columns 1-5), or jointly for that individual's decisions on all 5 MCLT tasks (Column 6). Actual bias $B^A_{i,m}$ is calculated as the difference between the naive $\tilde{\theta}^N_{i,m}$ 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 our model with endogenous effort based on all 55 binary choice tasks. Bias predicted by our model is calculated as the difference between the biased $\theta^N_{i,m}$ predicted by our model to be implied by individual i's choice under insufficient effort on multiple choice task m (or, in the last column, as an average implied by his choices on all 5 multiple choice tasks) and $\theta^*_{i,m}$ that would be inferred if the individual put in sufficient effort to choose his preferred option on multiple-choice task m.

mula derived from our theoretical model is effective in removing bias in risk aversion implied by observed choices without requiring any further estimation.

7 Reconciliation of Competing Discrete Choice Models

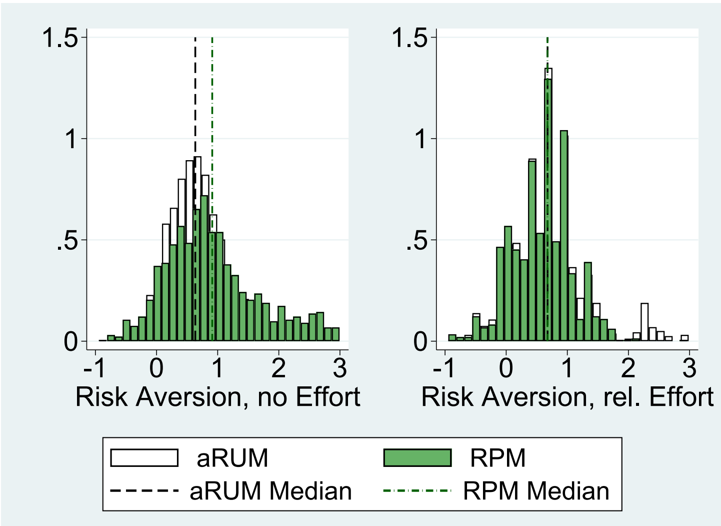
In the traditional Random Utility Model with additive i.i.d shocks (aRUM), the error term is appended to an individual's utility. Apesteguia and Ballester (2018) show that the aRUM as traditionally specified is not monotone when applied to risk preferences. Intuitively, the likelihood of preferring the 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 utility 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 $\sigma^{RU}_{\theta,i}$ 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 in case of 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 shifted to the right when using preferences shocks rather than additive utility shocks (see the left histogram of Figure 8 below).[68]

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



Notes: The figures plot distributions of the coefficient of relative risk aversion in the sample, estimated alternatively using an aRUM model with additive utility shocks (transparent bars) and a RPM model with preference shocks (green bars). The left panel omits endogenous relative effort from estimation while the right panel includes it.

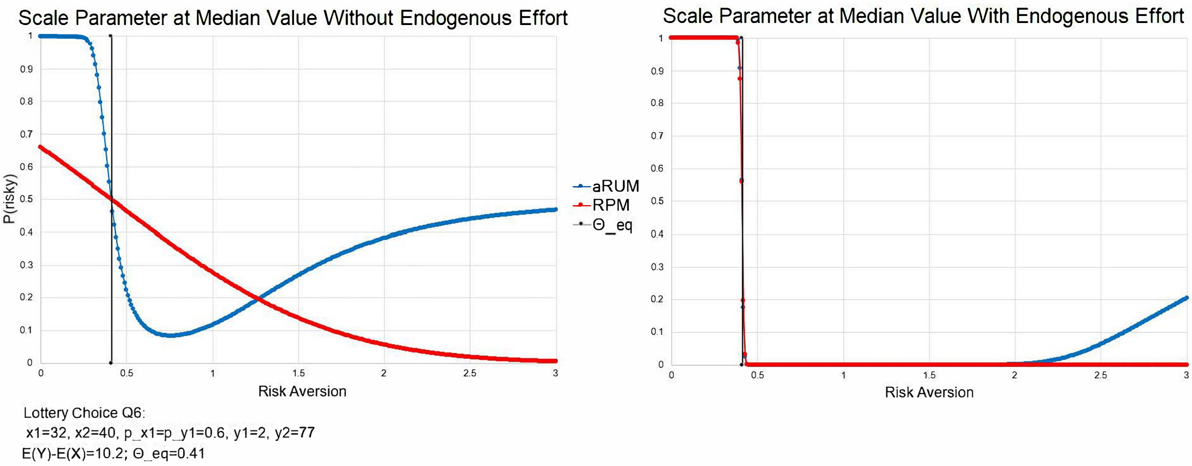
Once we estimate our model with endogenous effort, the non-monotonicity of the aRUM becomes empirically irrelevant, at least in the context of our experimental sample. The distributions of the coefficient of relative risk aversion estimated using either preference shocks or additive utility shocks converge (see the right histogram of Figure 8), with differences only visible at high levels of risk aversion which are uncommon in our sample. 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.[69] After taking into account endogenous effort, one could thus put

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

[69] This is due to the fact that we are combining a non-monotone choice model (aRUM) with a quasi monotone one

risk preference estimation within an aRUM framework in the same category as time preference estimation: problematic in theory but less so in practice.[70]

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



Notes: The figure illustrates the non-monotonicity problem of the aRUM, using one of the lottery choice tasks included in the experiment. The probability of choosing the riskier of two options is plotted against the coefficient of relative risk aversion. The blue line represents the prediction under an aRUM model with additive utility shocks while the red line represents the prediction under a RPM model with preference shocks. The threshold level of indifference between the two lotteries is $\theta^{eq} = 0.41$. The error shock is always mean zero. The left panel uses the estimated standard deviation of the error shock for the median individual when endogenous relative effort *is not* accounted for, while the right panel uses the estimated standard deviation of the error shock for the median individual when endogenous relative effort *is* accounted for. The standard deviation estimates are from the respective models (aRUM or RPM) used to obtain the choice probability.

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 relative 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

(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.

[70] 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.

problem of the aRUM. The graph on the right assumes error shocks of a magnitude estimated for the median individual *when the relative 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 ($\theta_i > 2$ using the example task).[71] 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 and aRUM once we allow the decision to exert effort to depend on an individual's perceived costs and benefits of doing so.

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 at the provided incentive level. 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 an individual's latent risk preference).

Omitting the initial effort decision results in estimates of risk aversion biased by approximately 50% for the median individual on the more complex tasks. Overestimation of risk aversion 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

[71] Correspondingly, at high values of risk aversion, we see imperfect convergence of the estimated distributions of the coefficient of relative risk aversion using the aRUM and RPM once endogenous relative effort is taken into account.

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. The statistically significant differences in how men and women respond to the costs and benefits of exerting effort suggests a nuanced pattern of bias in preference estimates for different demographic groups which merits further exploration.

Our results illustrate that a sophisticated model of decision noise is much less important on tasks where individuals find it worthwhile to pay sufficient attention to choose according to their true preference, given the available incentives, and choices are thus largely uncontaminated by decision noise. It appears that the simpler choice design used in this experiment fits that description pretty well. Simple and complex models of behavior thus yield similar 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.[72] 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 thus 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 effort required to choose according to latent preferences on a given task is thus sufficiently low such that most individuals find 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

[72] 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.

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 the 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 that decision noise may interact with an experimental elicitation design to produce upwards or downwards bias in preference estimates (risk, time, social, etc.), which manifests itself as apparent preference instability when not taken into account. We apply the model 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 use our model to derive a simple formula for the bias and demonstrate that it generalizes to a related out-of-sample setting involving incentivized choices between multiple lotteries.

We find that 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 essentially invariant to elicitation task complexity. Indeed, preferences are essentially 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 relative effort is predictive of an individual's high school GPA and performance on an achievement test used to make international comparisons. This suggests that it captures an individual's propensity to exert effort which generalizes to other settings. 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).

Our results suggest that accounting for decision noise which systematically varies with task attributes is a fruitful avenue for obtaining better estimates of true latent preferences from observed decisions. In this paper, we explored shifters in the costs and benefits of effort; however, the set of potential determinants of decision noise is much wider. Future research should aim to disentangle the impact of an expanded set of environmental and task design features on the decision noise content of observed choices. We see this as a path towards explaining the seemingly incongruous preference estimates when a feature of the decision environment is altered, which currently generate much attention in the literature. As more systematic variation is accounted for, modeling the residual randomness in decisions as classic white noise will be more plausible. 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. 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. Finally, the importance of low-stakes and high-stakes motivation in real-world settings also merits further study.

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.” *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.” *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.* Working Paper.

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

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, T., A. Falk, D. Huffman, and U. Sunde. 2010. “Are Risk Aversion and Impatience Related to Cognitive Ability?” *American Economic Review* 100 (3): 1238–60.

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.

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.

Gonçalves, D. 2024. "Speed, Accuracy, and Complexity." ArXiv preprint arXiv:2403.11240. arXiv: 2403.11240. https://arxiv.org/abs/2403.11240.

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." *Netspar Industry Series Design Paper 233.*

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.

Keffert, H., and N. Schweizer. 2024. “Stochastic Monotonicity and Random Utility Models: The Good and The Ugly.” *arXiv preprint* arXiv:2409.00704. https://arxiv.org/abs/2409.00704.

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.”

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.

OECD. 2010. *PISA 2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science (Volume I).* PISA. Paris: OECD Publishing. https://doi.org/10.1787/9789264091450-en.

OECD. 2019. *PISA 2018 Assessment and Analytical Framework.* PISA. Paris: OECD Publishing. https://doi.org/10.1787/b25efab8-en.

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.

Steiner, Johannes, Christopher Stewart, and Filip Matějka. 2017. “Rational Inattention Dynamics: Inertia and Delay in Decision-making.” *Econometrica* 85 (2): 521–553.

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.

Vieider, F. M. 2024. “Decisions under uncertainty as bayesian inference on choice options.” *Management Science* 70 (12).

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

A.a Appendix Tables

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: Representative Agent Model: Raw Coefficient Estimates

	(1)	(2)	(3)	(4)
Coefficient of Relative Risk Aversion	0.67***	0.67***	1.10***	0.68***
	(0.01)	(0.01)	(0.03)	(0.01)
Cognitive Noise Intercept	-0.63***	-0.64***	0.61***	-0.67***
	(0.02)	(0.02)	(0.02)	(0.02)
Relative Effort Intercept	2.13***	2.00***		1.46***
	(0.11)	(0.15)		(0.14)
Relative Effort: Stakes	-0.02	0.02		0.25***
	(0.03)	(0.04)		(0.08)
Relative Effort: Fatigue	-0.23***	-0.23***		-0.28***
	(0.03)	(0.03)		(0.04)
Relative Effort: Complexity	-2.04***	-1.89***		-1.26***
	(0.13)	(0.18)		(0.17)
Relative Effort: Intercept*Sex				1.09**
				(0.46)
Relative Effort: Stakes*Sex				-0.42***
				(0.13)
Relative Effort: Fatigue*Sex				0.12**
				(0.06)
Relative Effort: Complexity*Sex				-1.30**
				(0.53)
Number of observations	67,320	63,648	63,648	63,648
Dominated choices excluded		x	x	x
Log-likelihood	31,398	31,294	36,130	31,261
Akaike Information Criterion	62,807	62,601	72,264	62,542
Schwartz Information Criterion	62,862	62,655	72,282	62,632

Standard errors clustered at the individual level in parentheses
*** $p < 0.01$, ** $p < 0.05$
Notes: In order to allow for comparability with the model which omits the relative effort decision, columns (2)-(4) exclude 3 choice tasks per individual which feature a dominated option. To accommodate the model without relative effort, the standard deviation of the preference shock is allowed to be unbounded, as, in that specification, it not only reflects pure cognitive noise but also picks up on other sources of inconsistency in individuals' decisions. It can be calculated as exp(Cognitive Noise Intercept).

Table A.3: Representative Agent Model: Marginal Effects

	Men	Women
Baseline Relative Effort	0.81***	0.92***
	(0.04)	(0.08)
Marginal Effects		
Stakes	0.07***	-0.03
	(0.02)	(0.02)
Fatigue	-0.08***	-0.03**
	(0.01)	(0.01)
Complexity	-0.34***	-0.39***
	(0.05)	(0.08)

Standard errors clustered at the individual level in parentheses
*** $p < 0.01$, ** $p < 0.05$
Results come from a representative agent model allowing for interactions between the components of endogenous relative effort decisions and the decision-maker's sex. Marginal effects denote the effect of a 1 standard deviation increase in the relevant relative effort shifter or, in the case of complexity, the effect of a change from the simpler task design to the more complex one. Estimates are calculated at average values of shifters of relative effort.

Table A.4: **Predictive Power of an Individual's Latent Risk Aversion Estimated from Binary Choices**: Explaining the Naive and De-Biased Level of Risk Aversion Implied by Choices Averaged Across all 5 MC Tasks

	Multiple Choice Average	
	Implied Naive $\tilde{\theta}^N_{i,M}$	De-Biased $\theta_{i,M}$
Endogenous Relative Effort Model $\hat{\theta}_i$		
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

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

Notes: Each column displays the results of an OLS regression. The dependent variable is the coefficient of relative risk aversion implied by an individual's choices on the 5 lottery multiple choice tasks. The explanatory variable is an estimate of that individual's risk aversion from binary lottery choice tasks. The "Endogenous Relative Effort Model $\hat{\theta}_i$" is obtained by estimating our model with endogenous effort using individual i's choices on all 55 binary choice tasks. The naive $\tilde{\theta}^N_{i,M}$ is calculated from individual i's choices on 5 multiple choice tasks, using indifference thresholds associated with the constituent lotteries. The de-biased $\theta_{i,M}$ is obtained by applying the bias correction implied by Equation 18 to the naive $\tilde{\theta}^N_{i,M}$.

A.b Adapting the Model with Endogenous Relative Effort to Inference of Risk Aversion in a Multiple-Choice Lottery Setting

We proceed in three steps:

1. We first calculate the *naive* coefficient of relative risk aversion $\tilde{\theta}^N_{i,m}$ implied by individual i's observed choice on MCLT task m.[73] As we perform no estimation here, we simply take the average of the 2 indifference thresholds around a chosen lottery on task m to obtain the relevant $\tilde{\theta}^N_{i,m}$.[74]

   We then calculate the implied coefficient of relative risk aversion $\tilde{\theta}^N_{i,M}$ that would be inferred from all 5 individual i's observed multiple choice decisions as a simple average of the $\tilde{\theta}^N_{i,m}$ implied by each of the 5 observed choices:

$$\tilde{\theta}^N_{i,M} = \frac{\sum_{m=1}^{M} \tilde{\theta}^N_{i,m}}{M} \tag{17}$$

[73] The calculated risk aversion indifference thresholds for the 5 ordered groups of binary cMPL choice tasks give the respective thresholds for the 5 MCLT tasks (see Table B.2 of the Appendix). This is illustrated in Figure 7 for the first multiple choice lottery task which combines the 5 cMPL tasks from Figure 1 into a single task. The indifference thresholds represent the level of risk aversion at which an individual would be indifferent between two *adjacent* lotteries in a given MCLT task. Individuals with a θ_i above the highest indifference threshold in a given MCLT task will prefer the safe lottery. Individuals with a θ_i below the lowest indifference threshold in a given MCLT 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. The assumption is reasonable given our finding that once effort is taken into account, the scale of the error shock tends towards 0 for the median individual.

[74] 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 MCLT task) or *subtract* it (if he chose the riskiest lottery).

According to our model, we expect $\tilde{\theta}^N_{i,M}$ to be biased for individuals who do not put in sufficient effort to make a choice in line with their latent risk preference on the MCLT tasks.

2. For each of the five MCLT tasks, we next determine an individual's *preferred lottery* based on our estimate of that individual's latent coefficient of relative risk aversion $\hat{\theta}_i$, obtained in Section 5.b by applying our model with endogenous relative effort to all 55 observed binary lottery choices of individual i.[75] We then calculate the implied coefficient of relative risk aversion $\theta^*_{i,m}$ that would be inferred from an individual's choice of his preferred lottery 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 $\theta^*_{i,M}$ that would be inferred if individual i chose his *preferred lottery* on all 5 MCLT tasks, by averaging the constituent $\theta^*_{i,m}$, analogously to Equation 17.

3. Finally, for each individual i and MCLT task m, we can now calculate the predicted level of bias in the naive $\tilde{\theta}^N_{i,m}$, given the characteristics of task m, individual i's estimated true latent risk aversion $\hat{\theta}_i$, and his estimated average relative effort on the more complex binary tasks, which we denote $\hat{E}^R_{i,cMPL}$.[76] We obtain it as the difference between the biased naive $\theta^N_{i,m}$ that our model *predicts* we can expect to be implied by individual i's choice on task m given his estimated relative effort on similar binary tasks, and $\theta^*_{i,m}$ which would have been obtained from the individual's choice of his truly preferred option on task m under sufficient relative effort. According to our model summarized in Equation 1, $\theta^N_{i,m}$ will be a weighted average between $\theta^*_{i,m}$ (chosen when the individual exerts sufficient relative effort, so $\hat{E}^R_{i,cMPL}$ 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 to exert no effort). More precisely, our model predicts that the bias $B^M_{i,m}$ in the naive coefficient of relative risk aversion, for individual i based on his choice on MCLT task m with z options is:

$$B^M_{i,m} = E(\tilde{\theta}^N_{i,m} - \hat{\theta}_i) = \theta^N_{i,m} - \theta^*_{i,m} \tag{18}$$

   with

$$E(\tilde{\theta}^N_{i,m}) = \theta^N_{i,m} = E^R_{i,m} \cdot \theta^*_{i,m} + (1 - E^R_{i,m}) \cdot \frac{\sum_{z=1}^{Z} \theta_{z,m}}{Z} \tag{19}$$

[75] 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 estimation by applying our model to individual i's observed binary choices–and values of the coefficient of relative risk aversion $\tilde{\theta}^N_{i,m}$ implied by the individual's multiple choice data and calculated separately without the use of any statistics or econometrics. Our model will have out-of-sample predictive power if $\hat{\theta}_i$ predicts $\tilde{\theta}^N_{i,m}$.

[76] $\hat{E}^R_{i,cMPL}$ can be seen as the *upper* bound on individual i's propensity to exert sufficient relative effort on the MCLT tasks as these are even more complex than tasks of the binary cMPL design, while having on average the same stakes as the cMPL binary tasks and involving the same (or greater) mental fatigue, because the MCLT 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.

where *E(.)* is the expectation operator, $\theta^N_{i,m}$ is the naive (biased) coefficient of relative risk aversion that our model *predicts* an analyst can expect to infer for individual i on task m given the characteristics of task m, individual i's estimated true latent risk aversion $\hat{\theta}_i$, and $E^R_{i,m}$: his relative effort exerted on task m; $\tilde{\theta}^N_{i,m}$ is the naive coefficient of relative risk aversion inferred from individual i's *actual choice* on task m using the indifference threshold method outlined above, $\theta^*_{i,m}$ is the coefficient of relative risk aversion that would be inferred using the indifference threshold method if individual i chose his preferred (expected utility maximizing) option on task m given his $\hat{\theta}_i$, z is the number of alternatives that the individual is choosing between on task m, and $\theta_{z,m}$ is the coefficient of relative risk aversion that would be inferred from a choice of lottery z on multiple choice task m using the indifference threshold method. For the purposes of the calculation, we assume that individual i has a constant propensity to exert sufficient effort across the MCLT tasks, equal to his average estimated relative effort on the more complex cMPL binary choice tasks, so $E^R_{i,m} = \hat{E}^R_{i,cMPL}$. Equation 19 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.[77]

The bias correction is obtained by subtracting predicted bias implied by Equation 18 from the naive coefficient of relative risk aversion: the de-biased $\tilde{\theta}_{i,m} = \tilde{\theta}^N_{i,m} - B^M_{i,m}$.

[77]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 MCLT task m and the individual's coefficient of relative risk aversion estimated based on his 55 observed decisions on binary choice tasks, so $B^A_{i,m} = \tilde{\theta}^N_{i,m} - \hat{\theta}_i$.