JOB MARKET PAPER

Separating Preferences, Skills, and other Latent Personal Attributes from Endogenous Effort and Cognitive Noise*

Please click here for the latest version.

Author:

Christian BELZIL† and Tomáš JAGELKA‡

November 14, 2023

*We would like to thank Jose Apesteguia, Miguel Ballester, James Heckman, Erzo F.P. Luttmer, Antonio Penta, Douglas Staiger, Robert Sugden, John Rust, and Erik Wengström for their helpful comments and suggestions. We also thank participants at the CIRANO workshop in Behavioral and Experimental Economics for Innovative Policy Making, the European Economic Association Congress, the ESA Junior Faculty Webinar, ESA Global Meetings, at the Economic University of Bratislava, at the IAREP Early Career Workshop, and at the Czech Economic Society and Slovak Economic Association Meeting for valuable feedback. This research is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC 2126/1– 390838866.

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

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

Abstract

Preferences, skills, and other latent personal attributes (PSAs) are key drivers of inequalities in life outcomes. We propose a novel framework for quantifying, and accounting for, individuals' effort and cognitive noise which confound estimates of PSAs based on observed behavior. We establish the ability of our framework to quantify the noise content of a given experimental design and to de-bias estimates of PSAs, in an application to a large-scale experimental dataset measuring risk preferences. While the two elicitation designs we study were used interchangeably in the past, we estimate that a change from the more complex design to the more intuitive one results in a 30% decrease in (rational) inattention. On the one hand, failure to properly account for decision errors results in estimates of risk preferences biased by 50% for the median individual. On the other hand, accounting for endogenous effort allows us to empirically reconcile competing models of discrete choice. Furthermore, the estimated individual effort propensities have external validity. We show that they capture low-stakes motivation which generalizes to other settings and predicts, *inter alia*, an individual's performance on the highly influential PISA achievement test.

1 Introduction

There is much inequality in essential life outcomes such as education, earnings, and longevity. Initially, economists attempted to explain differences in economic well-being primarily by differences in the amount of physical capital. Becker's (1964) seminal contribution was to turn the focus to *human capital*. Over the past half-century, much effort has been put into understanding the number and nature of its constituent attributes. Insights from psychology and neuroscience enrich economic models of behavior, which increasingly recognize that heterogeneous preferences interact with cognitive and non-cognitive skills, decision-making ability, motivation, information, behavioral biases, and resource constraints to produce outcomes (e.g., Heckman, Stixrud, and Urzua, 2006; Guadalupe, 2007; Almlund et al., 2011; Heidhues and Kőszegi, 2017; Deming, 2017; Falk et al., 2018; Coffman and Klinowski, 2020; Todd and Zhang, 2020; Heckman, Jagelka, and Kautz, 2021). This introduces a pressing demand for accurate measurement of preferences, skills, and other latent personal attributes (henceforth "PSAs"): researchers need to refine and empirically validate theoretical models; employers seek to understand bundles of skills and attributes which maximize employees' productivity; and policymakers want to understand the preferences of the governed in order to design policy which aligns incentives with those preferences.

Unobserved PSAs need to be inferred from observed behavior. In order to elicit any particular PSA, we need to isolate its impact on observed choices from confounders. Heckman, Jagelka, and Kautz (2021) clarify that performance on *any* task (including task performance in controlled settings such as surveys or experiments) is a function of multiple preferences, skills, and also of effort, which in turn depends on task-specific incentives. Our main contribution is to develop a **sequential stochastic choice** model which separates **endogenous effort** from **cognitive noise**[1] due to imperfect self knowledge, and accounts for the effect of both these confounders on observed choices. By incorporating the (implicit) effort decision which necessarily precedes any choice, we provide a tool for quantifying effort propensities across individuals, across task designs, and across task design features. We thus provide a framework for uncovering mechanisms underlying endogenous effort. By accounting for endogenous effort decisions, we are able to cor-

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

rect bias in estimates caused by mistakes due to inattention. By modelling the choice decision itself, which includes cognitive noise in the form of an error shock, we lay the groundwork for the reconciliation of competing models of discrete choice.

In our framework, a person first decides whether a task is worth paying attention to. Rather than assuming an exogenous mistake probability, we allow this decision to depend on the perceived costs and benefits of exerting effort to make a choice.[2] Our model therefore treats the reliability of each decision as the result of an endogenous process, a form of rational inattention.

If the perceived benefits of effort exceed the costs, the individual subsequently chooses the option which provides the highest expected utility, given task attributes and a latent PSA of interest. Otherwise, the individual activates effortless System 1 thinking, described by Kahneman (2011), which does not require the evaluation of expected utilities of the proposed choice alternatives, and answers using a heuristic or randomization strategy.

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

A natural application of our endogenous effort model is to the estimation of economic preferences. We illustrate the usefulness of the model for quantifying the impact of various experimental design features on the signal-to-noise ratio of responses using the example of risk preferences.

[2]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.

[3]This is evidenced by frequent inconsistent choices on repeated tasks even in controlled laboratory settings (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) and by test-retest correlations well below the noise-free benchmark of "1" for repeated survey measurements elicited on the same sample within a short enough time period (e.g., a few weeks) such that the underlying attributes of interest can reasonably be assumed stable (e.g., Krueger and Schkade, 2008; Lang et al., 2011; Beauchamp, Cesarini, and Johannesson, 2017; Soto and John, 2017; Falk, Neuber, and Strack, 2021; Dohmen and Jagelka, forthcoming).

In standard economic models, risk preference is embodied by the coefficient of risk aversion and determines the shape of an individual's utility function. Experimental economists use incentivized choice tasks to elicit risk preferences in controlled laboratory settings. Multiple task designs exist, which are often used interchangeably, with no systematic understanding of the impact of design variations on the measurement properties of an instrument. Our structural model allows us to endogenize the decision to exert effort on a given experimental task, as a function of the task's characteristics and of unobserved individual heterogeneity. We use it to show that individuals' responses on two popular designs used in the literature for eliciting risk preferences contain widely varying amounts of noise which bias estimates of risk aversion by 50% for the median individual when not properly accounted for.

We estimate the model using a representative sample of 1,224 Canadian high school seniors, each of whom made choices on 55 incentivized tasks used to elicit risk preferences. There are two types of such choice tasks in this experiment. Both use the Multiple Price List (MPL) setup which relies on ordered groups of binary choice tasks between lotteries. The relative attractiveness of the riskier lottery is monotonically increasing or decreasing within an MPL. The HL design (as in Holt and Laury, 2002) uses variations in probabilities along with fixed payoffs within an MPL while the Ordered Lottery Selection - "OLS" - design (as in Eckel and Grossman, 2008) uses variations in payoffs with fixed probabilities.[4] In this experiment, each individual makes choices on three MPLs of the HL design, each consisting of ten choice tasks, and five MPLs of a binarized OLS design, each consisting of five choice tasks. We describe them in more detail in Section 4.

Within each MPL of the HL design, 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 OLS design where this pattern is not easily discernible. One might thus expect more mistakes and more noise on the OLS design due to rational inattention. We quantify this intuition.

We find that voluntary mistakes (those driven by low effort) increase with task complexity, with low relative stakes, and with fatigue—instances in which costs of making the right choice are higher and benefits lower. Changing the task design from OLS to HL by itself results in a 30% increase in the likelihood of exerting sufficient effort for the median individual. Accordingly, while underlying risk preference accounts for 90% of explained cross-sectional variation in lottery choices of the HL design, almost half of explained cross-sectional

[4] Harrison and Elisabet Rutström (2008) provide an excellent summary on the various experimental designs and techniques used to elicit risk preferences in the laboratory.

variation is attributed to noise due to inattention on the OLS design.

Accounting for endogenous effort is particularly crucial when observed choices contain a lot of noise. While the distribution of the coefficients of risk aversion based on HL tasks is largely unchanged if endogenous effort is omitted, omitting effort on the OLS design biases risk aversion estimates by 50% for the median individual. As predicted, the bias is higher for individuals who have a high likelihood of making mistakes and for whom the experimental design is particularly lopsided towards the riskier alternative given their true risk preference. This quantifies Andersson et al. (2016, 2020)'s claim that the interaction of random decision errors with an experimental design and an individual's latent preferences may introduce complex patterns of bias in preference estimates, when sources of noise are not properly accounted for.

Despite strong evidence in favor of HL being the cleaner design for measuring risk preferences, in the context of this experiment the two types of tasks are complementary. Excluding OLS tasks which cover a wider range of risk aversion results in a significant loss of explanatory power for individuals with an estimated coefficient of risk aversion greater than +2. This underscores the importance of correctly accounting for the noise due to endogenous effort inherent in OLS tasks, as excluding them completely may be suboptimal.

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. Observed choices match those predicted in simulations approximately 90% of the time. Structural estimates explain choices on OLS tasks less well than on HL tasks consistent with a bigger role of noise in decisions on the former.

We show that our estimated effort propensity also has **external validity** and is particularly predictive of an individual's performance in low-stakes environments. We thus call it *low stakes motivation*. A one standard deviation increase in low-stakes motivation could affect the PISA numeracy ranking of a mid-performing country by approximately 9 places (a 40% jump in the rankings). Furthermore, we provide evidence that the propensity to exert effort in low stakes settings is fundamentally different from the propensity to exert effort in high stakes settings.

Even when individuals exert sufficient effort, residual randomness in individuals' choices from the point of view of the econometrician often remains (e.g., Dohmen and Jagelka, forthcoming). Such cognitive noise can be modeled as shocks to utility. A controversy arose recently in the literature as to where the stochastic shock should be placed when modeling economic preferences. In the standard additive Random Utility Model ("aRUM"), *the shock is appended to utility*. In a Random Preference Model ("RPM"), *the shock enters utility via preferences*. Apesteguia and Ballester (2018) prove that under standard assumptions on the utility function, the aRUM, unlike the RPM, exhibits anomalies in predicted choice probabilities under risk (and intertemporal delay) which call into question its continued use in preference estimation.

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

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

The rest of the paper is organized as follows: Section 2 surveys the literature on random choice models, Section 3 presents the general structural model, Section 4 describes the data, Section 5 introduces the empirical application, Section 6 presents the main empirical results, and Section 7 provides a general discussion of the broader implications of our findings and concludes.

2 Background on Random Choice Models

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

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

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

Although the RPM is intrinsically monotonic, it leaves no room for processing error. Unlike the aRUM, it cannot explain lapses in attention which may cause some individuals to choose dominated choices.[5] The most common solution to this problem implemented in the experimental literature is to introduce a "tremble parameter" which captures the probability that an individual makes mistakes (Harless and Camerer, 1994). This essentially assumes that everyone evaluates the expected utility of each alternative and mistakes in decisions are purely random. Apesteguia and Ballester (2018) use a tremble parameter assumed to be common to the whole population while Jagelka (forthcoming) allows it to depend on both observed and unobserved heterogeneity. He shows that heterogeneity in the stability of preferences is separately identifiable from heterogeneity in individuals' propensity to make mistakes and that these two components of random decision-making can be linked to separate non-cognitive and cognitive skills.

Additive random utility shocks or tremble parameters can be regarded as involuntary (exogenous) mistakes. Interpreting all mistakes as involuntary may however be unrealistic. When individuals see the choice tasks as relatively complex or perceive no meaningful difference be-

[5] 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.

tween lotteries, they may judge that the disutility cost of solving the expected utility problem is too high compared with potential benefits. For this reason, we endogenize the decision to pay attention.

3 Model

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

When an individual is presented with a choice task, he first examines the readily and effortlessly available characteristics of the options among which he has to choose and decides whether or not it is worth to expend effort on making the choice. If it is, the individual exerts the amount of effort necessary to choose according to expected utility maximization given his relevant latent PSA. If it is not, effortless System 1 thinking described by Kahneman (2011) is activated and the individual answers according to some heuristic or randomization strategy.

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

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

where $P(YC_{i,l}=1)$ is the probability that individual i *chooses* option Y on choice task l; $P(YP_{i,l} = 1)$ is the probability that individual i *prefers* option Y on choice task l; $P(E_{i,l}=1)$ is the probability that individual i will choose to exert effort on choice task l; and $p_{Y,i}$ is the probability with which individual i picks option Y when he chooses not to exert effort. It can be understood as his "effortless" randomization strategy or heuristic. A reasonable default value is $p_{Y,i} = 0.5$, i. e., in the absence of effort, an individual randomizes between the available options with equal probability.

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

3.a Decision to Exert Effort

Each individual first briefly "takes in" a task, noticing its *readily and effortlessly available* characteristics. For the purposes of this exercise, we only consider such characteristics which pertain to the perceived costs and benefits of exerting *sufficient effort* to pick the preferred alternative on a given choice task.[6] Denote C_l the vector of readily and effortlessly available characteristics of choice task l which pertain to the perceived *costs* of exerting sufficient effort, and denote B_l the vector of readily and effortlessly available characteristics which pertain to the perceived *benefits* of exerting sufficient effort. Let us assume that the individual acts according to *net* perceived benefits and that effort is indivisible, i. e., conditional on choosing to exert effort, the individual will exert *sufficient* effort for making a choice according to his latent preference.

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

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

where $b_{0,i}$ is the intercept, $b_{1,i}$ 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, $\varepsilon^b_{i,l} \in (0;\infty)$ is an i.i.d. random shock realized before an individual considers the net benefits of exerting effort on choice task l, and $cdf(\cdot)$ denotes the relevant cumulative distribution function given the distribution for $\varepsilon^b_{i,l}$.

3.b Preference Between Available Options

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

When an individual is faced with a choice between two options X and Y—in a deterministic world

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

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, such as imperfect self-knowledge, need to be considered (e.g., Loomes and Sugden, 1995; Kahneman, 2011; Enke and Graeber, 2023; Jagelka, forthcoming).[7] Indeed, even when individuals exert sufficient effort, residual randomness in individuals' choices from the point of view of the econometrician often remains (e.g., Dohmen and Jagelka, forthcoming).

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

When an individual is faced with a choice between two options in the presence of utility shocks, even conditional on exerting sufficient effort his preference over the options will be probabilistic unless one option is dominated by the other, i. e., there is no value of the error shock which would make it the preferred option. Without loss of generality, option Y is preferred when $DEU_i(m_y, n_y, o_y; \Psi_i; \varepsilon) > DEU_i(m_x, n_x, o_x; \Psi_i; \varepsilon)$.[8] 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 pre-

[7] Alternatively, individuals may simply have a *preference* for randomization (see Agranov and Ortoleva, 2017).

[8] In full, option Y is preferred when $DEU_i(Y; \Psi_i; \varepsilon_y) > DEU_i(X; \Psi_i; \varepsilon_x)$. When ε directly affects a preference parameter, $\varepsilon_x = \varepsilon_y = \varepsilon$. When ε is an additive utility shock, we can always combine the shocks to obtain $\varepsilon = \varepsilon_y - \varepsilon_x$ because *differences* in utility determine the preferred choice.

ferred, the effort decision determines whether an individual converts the preference into an actual choice.

4 Data

We illustrate the application of our model from Section 3 using experimental data designed to elicit risk preferences. An individual is faced with 55 choice tasks of two designs (HL and OLS) which elicit risk preferences. Each task consists of a binary choice between lotteries with different expected payoffs and payoff variances. Choices were incentivized and students were paid for one randomly drawn decision at the end of the session. The availability of a long panel makes this an ideal setting to study decision noise at an individual level. Each choice provides information about an individual's risk aversion parameter provided that he takes the task seriously. The characteristics of the lotteries that are readily and effortlessly available to each individual and therefore factor into the effort decision are: task design and ordering (costs) and choice stakes (benefits). The full experimental setup is included in the Online Appendix.

The data comes from "The Millennium Foundation Field Experiment on Education Financing", which involved a representative sample of 1,224 Canadian citizens who were full time students in their last year of high school. The students were between 16 and 18 years old 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.

Choice tasks of both the HL and the OLS type are designed to be clear and simple (require little ability), homogeneous (same situation), and incentive-compatible (providing an incentive for individuals to choose according to latent risk preference). "The key assumptions behind this set-up are that the individual understands probabilities and the expected values of options being offered, and that other factors that may affect risky choice besides latent preference (for example, wealth), can be controlled for adequately," (Dohmen et al., 2018). However, in reality these assumptions may not hold fully.

4.a Holt & Laury's (HL) Multiple Price List Design

Of the 55 tasks designed to measure risk aversion, the first 30 are of the Holt and Laury (HL) type introduced by Miller, Meyer, and Lanzetta (1969) and used in Holt and Laury (2002). Choice payments and probabilities are presented using an intuitive pie chart representation popularized by Hey and Orme (1994). There are three groups of ten questions each. 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. While the expected value of both lotteries increases, the riskier option becomes relatively *more* attractive, meaning that in each subsequent row, the cutoff value for risk aversion below which individuals prefer the riskier lottery is increasing. As in the first row of each set of questions the expected value of the safer lottery X is greater than that of the riskier lottery Y, all but risk-seeking individuals should choose the safer option. Midway through the ten questions, the expected value of the riskier lottery Y becomes greater than that of the safer lottery X. At this point, risk neutral subjects should switch from the safer to the riskier option. In the remaining rows the relative attractiveness of lottery Y steadily increases until it becomes the dominant choice in the last row.[9] By the last row of each set of HL questions, all individuals are expected to have switched to the riskier option. Each person's "switching point" should be indicative of his risk aversion. By design, in the absence of a shock to either preferences or utility, each individual should switch at exactly the same point on the 3 sets of HL questions.[10]

The HL design minimizes mental processing costs associated with making a choice. Initial choices in each MPL are simple for most individuals as the safer lottery also offers a higher expected value. The increasing attractiveness of the riskier option within each MPL is clearly visible. In a sense, individuals are being gently nudged towards the point at which they wish to switch from the safer to the riskier lottery. This makes it a very simple and intuitive setting to elicit preferences.

4.b Binswanger's Ordered Lottery Selection (OLS) design

The remaining 25 tasks designed to measure risk aversion used in this study are a binarized version of the ordered lottery selection (OLS) design developed by Binswanger (1980) and popularized by Eckel and Grossman (2002, 2008). A similar task design was used in Engle-Warnick, Laszlo, and Escobal (2006). They consist of five groups of five questions each. Once again, in each group of questions, subjects are presented with an ordered array of binary lottery choices. In each choice task, they choose between lottery X (safer) and lottery Y (riskier). This time, lot-

[9] In the last row of all three sets of HL-type 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.

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

tery X offers a certain amount in the first row and all other alternatives increase in expected payoffs but also in their variance. In each subsequent row the riskier option becomes relatively *less* attractive, meaning that in each subsequent row, the cutoff value for risk aversion below which individuals prefer the riskier lottery, is decreasing. Most individuals are thus expected to switch from the risky to the safe option at some point (assuming that they initially picked the risky option). However, a risk neutral individual should always at least weakly prefer the riskier alternative.

Once more, the "switching point" should be indicative of an individual's risk preference. By design and unlike in the three sets of HL MPLs, under standard assumptions on the utility function (e.g., CRRA, CARA) the switching point should vary among the five sets of OLS MPLs for a given individual even if he is paying full attention and consistently choosing according to his true (or average) latent risk preference. In the absence of stochastic shocks to utilities of preferences, the HL tasks should allow for the identification of an interval for an individual's risk aversion while the OLS tasks should permit the refinement of this interval. Furthermore, while the HL tasks focus on the most common range of risk preferences (up to a coefficient of risk aversion of 1.37 under CRRA utility), OLS tasks let us identify highly risk-averse individuals. The two types of task are thus complementary.

On the binarized OLS design, the decreasing attractiveness of the riskier lottery within each MPL is not obvious. The absence of an easily discernible pattern within a group of tasks increases the per-task mental processing costs associated with making a choice based on one's latent risk preference. In this context, we might expect choices to reflect a mix of signal from latent risk preference and noise due to inattention as more individuals may decide that the tasks are not worth the effort required to evaluate them correctly.

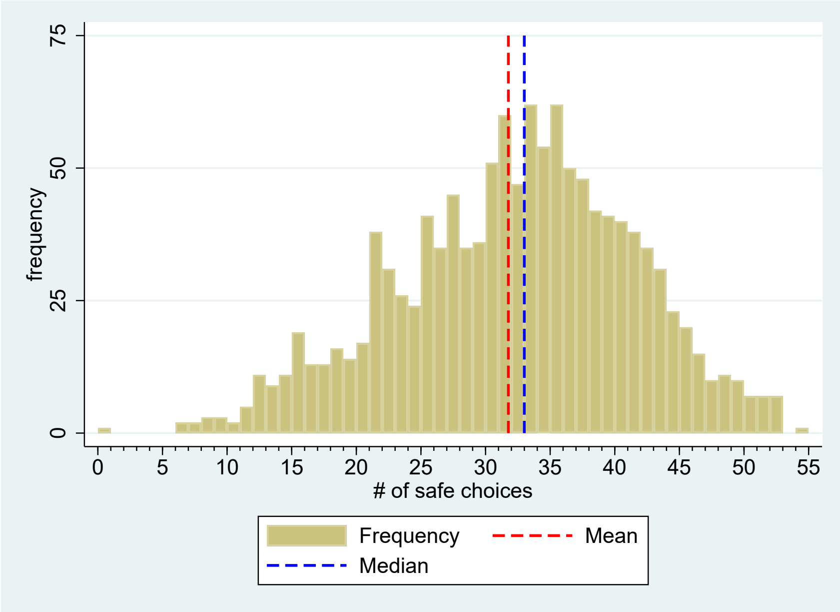
4.c Observed Individual Choices

Figure 1 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.

Contrary to standard predictions, many individuals exhibit reversals in their choices within a given MPL.[11] This shows the utility of collecting data on the full set of tasks as opposed to as-

[11] A reversal is defined as follows. Take for example one MPL of the HL design which includes ten binary choice tasks ordered by increasing relative attractiveness of the riskier lottery. If an individual starts out by picking the

Figure 1: Distribution of Individual Choices on Lottery Tasks



suming 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).

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 preference instability may suffice to explain this behavior, choice reversals *within* a given MPL are indicative of highly erratic decision-making which suggests inattention.[12] These distinct patterns of choice inconsistency help separately identify mistakes and preference instability as discussed in more detail in Section 5.a.ii and 6.a.iii.

Figure 2 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 HL-type tasks, almost all of the action takes place on OLS-type 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, on the one hand the HL design has features which minimize the per-task

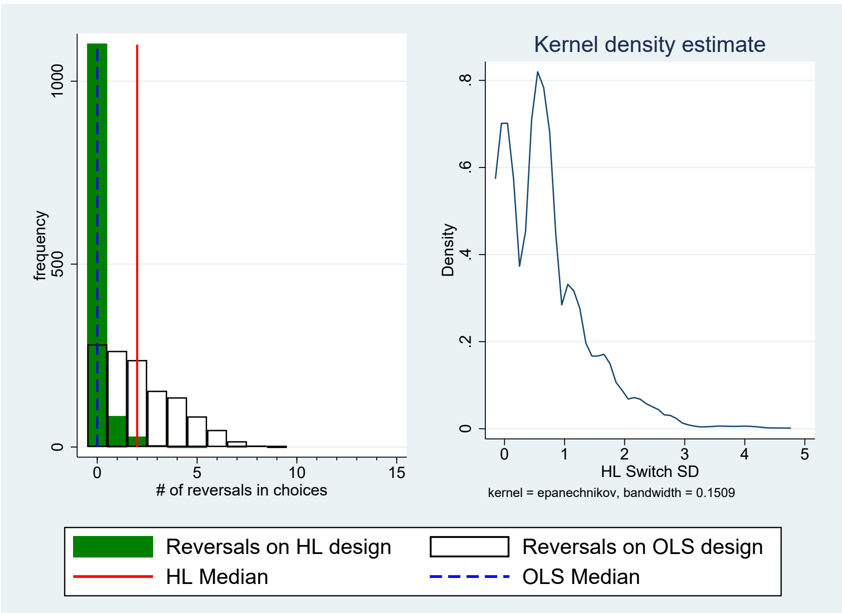
safer option and then at some point switches to the riskier one as the riskier option becomes more attractive, this is considered standard behavior. If however he then reverts back to the safer option on the same set of tasks even though the riskier option became *even more* attractive, this is considered a reversal. The definition is analogous for lottery tasks of the OLS design.

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

mental processing costs involved in choosing according to one's latent risk preference. On the other hand, choosing according to one's latent risk preference on tasks of the OLS design requires more mental effort. 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 (forthcoming) 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.

Inconsistencies in switching points can be easily detected on the three HL groups of tasks because they share common indifference thresholds under CRRA utility. We measure them as the standard deviation of switching points on the three MPLs of the HL design for each individual (0 implies consistent switching points across the HL tasks). The right graph of Figure 2 plots a distribution of switching point inconsistency on HL 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. A similar exercise cannot be done easily for the 5 groups of OLS tasks as predicted switching points on them differ. Our structural model is needed to detect such inconsistencies.

Figure 2: Observed Reversals per Individual on Lottery Choice Tasks



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 in Table A.1 .

Several recent papers analyze this dataset using a structural model. Jagelka (forthcoming) relates risk and time preferences to cognitive ability and psychological personality traits using a structural framework with observed and unobserved heterogeneity and a factor model. Belzil and Sidibé (2016) estimate individual preference over risk and time and study heterogeneity using various specifications of preferences, which include hyperbolic, quasi-hyperbolic discounting as well as subjective failure probability over future payments. They investigate the predictive power (transportability) of the estimated preference parameters. Belzil, Maurel, and Sidibé (2021) make use of the portion of the experiment devoted to preference elicitation in conjunction with information eliciting attitudes towards higher education to estimate the distribution of the value of financial aid for prospective students.

5 Application of the Endogenous Effort Model to Risk Preference Elicitation

The model described in Section 3.b is easily adapted to choice under risk: assume that (i) the effort cost vector C_l consists of a dummy for task design (allowing for task of the OLS design to have a different baseline effort requirement than task of the HL design, for example due to a different task complexity under each design) and an indicator for task order (allowing for the cost of a *given* amount of effort to change with fatigue); (ii) the effort benefit vector B_l consists of the percentage difference in the probability-weighted payoffs offered by each lottery (reflecting the stakes of each choice); (iii) if sufficient effort is exerted, an individual chooses according to expected utility maximization given his true coefficient of relative risk aversion and a *preference shock*, as in Jagelka (forthcoming), 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 is added directly on to the preference parameter; and (iv) if sufficient effort is not exerted, the individual randomizes between the two options with equal probability, i. e., $p_{Y,i} = 0.5$.

Let $U_i(a)$ represent the utility which an individual obtains from a dollars. Define the coefficient of relative risk aversion $\Theta_i = \frac{-a \cdot U''(a)}{U'(a)}$.[13] A CRRA utility function can then be written as:

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

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

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

If $\Theta_i \neq 1$,

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

If $\Theta_i = 1$,

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

When making a choice between lottery X and lottery Y, an individual first receives a realization of a preference shock, ε_i. We assume that the shock affects the individual's true (or average) risk preference embodied by his coefficient of relative risk aversion, Θ_i, which represents the relevant coefficient of risk aversion that would prevail in a purely deterministic choice context. In a stochastic choice environment, a random shock can reflect imperfect self-knowledge, actual variation in risk preference due to factors unobserved by the econometrician, or a preference for randomization. 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,...,55 that an individual faces. The individual will prefer the riskier lottery Y to the safer lottery X on task l if

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

The probability that Y is preferred on task l is equivalent to the probability that the value of the shock is such that the above inequality is satisfied. As $\varepsilon_{i,l}$ enters expected utility non-linearly, obtaining a closed-form expression for this probability is non-trivial. We use a trick provided by Apesteguia and Ballester (2018) to do so which relies on the monotonicity of the random preference model (RPM). For an in-depth discussion of an application of the RPM to the analyzed dataset, see Jagelka (forthcoming).

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.[14] Individual i will prefer the riskier lottery Y on task l if his shocked value of risk aversion is lower than the indifference threshold associated with task l:

$$\Theta_i + \varepsilon_{i,l} < \Theta_l^{eq} \tag{10}$$

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

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

Assuming that the random shock is normally distributed with $\varepsilon_{i,l} \sim N(0, \sigma_i^2)$, the probability that individual i prefers the riskier option Y on choice task l has a closed-form expression:[15]

$$P(YP_{i,l} = 1) = \Phi\left(\frac{\Theta_l^{eq} - \Theta_i}{\sigma_i}\right) \tag{12}$$

[14]Indifference thresholds for each of the 55 HL and OLS-type tasks in this experiment along with the percentage of individuals who picked the riskier option on each task are displayed in Tables A.6 and A.7. The three sets of choice tasks (MPLs) of the HL 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 OLS-type choice tasks are characterized by decreasing indifference thresholds reflecting a decreasing relative attractiveness of the riskier option. However, they do not exhibit the same congruence between the evolution of indifference thresholds and observed choices suggesting a more important role of noise on this task design and the need for a rich error specification in the structural model.

[15]Following Jagelka (forthcoming), we restrict $\varepsilon_{i,l}$ to plausible values, so $\varepsilon_{i,l} \in (0, 1]$.

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. We interpret Θ_i as the individual's true (or average) 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 uncertainty. The lower an individual's σ_i, the more consistent is his risk preference over a panel of choices he has to make.

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

5.a Identification of Consistency Parameters

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

5.a.i Identification Under Exogenous Effort

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

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

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

More generally, $p(E_i)$ and σ_i can be identified from different moments of the noise distribution, even in the absence of dominated choices. Essentially, there is a tension between the occurrence of inconsistent choices on task with a Θ_l^{eq} which is *close to*, or *far away from*, an individual's true (or average) risk preference Θ_i.[16] The resulting noise pattern is not sufficiently characterized by either consistency parameter alone.

5.a.ii Identification Under Endogenous Effort

Identification of endogenous effort parameters is more subtle than under exogenous effort, but follows the same general principles. The influence of readily and effortlessly available task characteristics, which factor into an individual's cost-benefit analysis of whether or not to exert effort sufficient for choosing according to his latent preferences, is identified from systematic differences in noise patterns for tasks with these characteristics. For example, take two task designs eliciting the same latent preference. If repeated choices on one of the designs are systematically more inconsistent/noisy than on the other design, one would expect to estimate a negative coefficient on that task design's influence towards an individual exerting sufficient effort. In our model, this would be interpreted as a higher cost of per-task effort required to choose according to latent risk preference on that task design.

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

[16] 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 true (or average) latent preference parameter.

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.

5.b aRUM Comparison

In the traditional Random Utility Model with additive i.i.d shocks (aRUM), the error term is appended to an individual's utility. For a choice between lottery X and lottery Y under aRUM, we thus have:

$$\begin{aligned} EU_i^{RU}(X) &= p_{x_1} \cdot U_i(x_1) + (1-p_{x_1}) \cdot U_i(x_2) + \varepsilon_{i,X}^{RU} \\ &= EU(X;\Theta_i) + \varepsilon_{i,X}^{RU} \end{aligned} \tag{14}$$

and

$$\begin{aligned} EU_i^{RU}(Y) &= p_{y_1} \cdot U_i(y_1) + (1-p_{y_1}) \cdot U_i(y_2) + \varepsilon_{i,Y}^{RU} \\ &= EU(Y;\Theta_i) + \varepsilon_{i,Y}^{RU} \end{aligned} \tag{15}$$

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

$$\begin{aligned} P(risky)_{i,l}^{RU} = P(YC_{i,l}=1)^{RU} &= P\left[EU_i^{RU}(Y) > EU_i^{RU}(X)\right] \\ &= P\left[\varepsilon_{i,Y}^{RU} - \varepsilon_{i,X}^{RU} > EU(X;\Theta_i) - EU(Y;\Theta_i)\right] \\ &= \Phi\left[\frac{EU(Y;\Theta_i) - EU(X;\Theta_i)}{\sigma_i^{RU}}\right] \end{aligned} \tag{16}$$

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

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 risk aversion Θ and the error scale parameter σ are identified if we have multiple binary choices between lotteries with varying payments and payment probabilities for each individual. As we have such information, we can estimate the aRUM model. 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 relevant to examine whether the non-monotonicity problem of the aRUM retains empirical relevance once endogenous effort is incorporated. See Section 6.d.

5.c Estimation

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

$$P(YC_{i,l} = yc_{i,l}) = P(YC_{i,l} = 1)^{YC_{i,l}} \cdot P(YC_{i,l} = 0)^{1-YC_{i,l}} \tag{17}$$

The likelihood contribution of individual i from all his observed choices is the probability of jointly observing his 55 lottery choices:

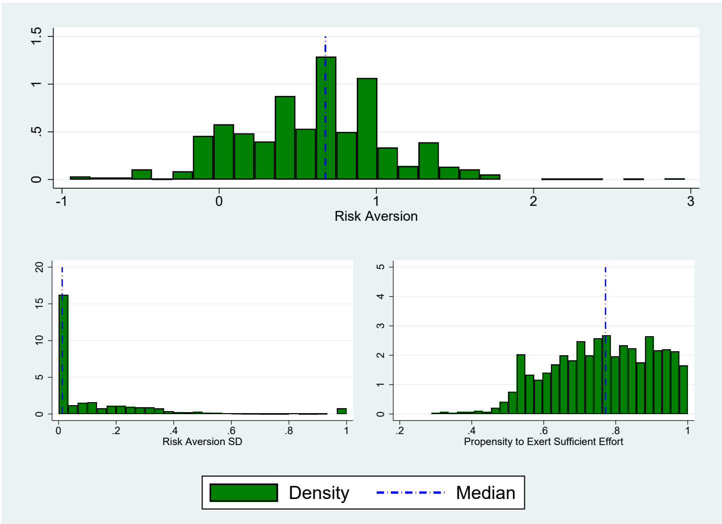
$$L_i = \prod_{l=1}^{55} P(YC_{i,l} = yc_{i,l}) \tag{18}$$

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

6 Empirical Results

Estimates from the full model with endogenous effort and cognitive noise based on observed choices on all 55 lottery tasks show that the median individual is risk averse, exhibits almost no cognitive noise, and exerts sufficient effort required for these tasks to give meaningful information about his latent risk preference about 75% of the time. The median and (mean) estimated values of the structural parameters are: 0.68 and (0.88) for the coefficient of relative risk aversion, 0.01 (0.13) for the standard deviation of the coefficient of risk aversion (a proxy for cognitive noise or imperfect self-knowledge), and 0.77 (0.76) for the propensity to exert sufficient effort for choosing according to underlying preferences, averaged over the 55 tasks that each individual faced. Figure 3 plots the parameter distributions.[17]

Figure 3: Distributions of Structural Parameters Estimated Using the Model with Endogenous Effort



In order to put these results in context, it is helpful to compare them to existing estimates. The obtained values of the coefficient of risk aversion are broadly in line with the previous literature (see e.g., Holt and Laury, 2002; Andersen et al., 2008; Apesteguia and Ballester, 2018; Apesteguia, Ballester, and Gutierrez, 2020; Jagelka, forthcoming).[18] 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, forthcoming), which is much higher than the value we obtain for the median individual.[19] Finally, I am not aware of any estimates of the percentage of the time that

[17]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.

[18]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.

[19]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

individuals put in sufficient effort on elicitation tasks, such that their choices reflect the underlying preference, skill, or some other latent personal attribute (PSA) that a researcher is trying to elicit.

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

6.a Endogenous Effort

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

The median individual is more likely to exert sufficient effort to choose according to latent preferences on HL than on OLS tasks, when stakes of getting the choice right are high, and when fatigue is low. The average impact of changing the design from OLS to HL is a 30% increase in the likelihood of exerting sufficient effort, P(E), for the median individual.[20] The marginal effect of increasing relative stakes by one standard deviation averaged across all 55 lottery choice tasks is a 7% increase in P(E) whereas increasing fatigue by one standard deviation results in a 2% decrease in P(E).[21]

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

between parameters using a Gaussian copula".

[20]This is consistent with the pattern of choice inconsistency observed in the raw data,which is concentrated on OLS tasks (see Figure 2).

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

6.a.i Determinants of Average Behavior

We find that our model fits the data well. We take key moments of the distribution of individual choices and regress them on the estimated structural parameters: the preference parameter Θ_i and consistency parameters σ_i and $P(E_i)$.[22] Row 2 of Table 1 shows that these explain over 80% of the cross-sectional variation in average choice behavior in terms of the percentage of the time that an individual selects the safer lottery and half of the variation in choice reversals. In comparison, the predictive power of demographic and socioeconomic variables is an order of magnitude smaller (see row 1 of Table 1).

Subsequent rows break down the explained variation in choices due to the estimated structural parameters into parts explained by the preference parameter and by the consistency parameters. This lets us compare their relative explanatory power, expressed as a percentage. Consistency parameters are further broken down into the standard deviation of risk aversion and the *propensity to exert effort*. This allows us to provide empirical evidence on the identification of the two types of consistency parameters based on different moments of choice inconsistency as outlined in Section 5.a.ii.

Almost 90% of the explained variation in observed choices is accounted for by the latent risk preference on HL tasks compared to only 50% on OLS tasks (the remainder is noise due to inattention or imperfect self-knowledge). Increasing the coefficient of risk aversion by one standard deviation leads to a 15% increase in the proportion of safe choices selected on HL tasks, compared to a 10% increase on OLS tasks.[23] This is yet another indicator that true risk preference has a larger impact on observed choices on the cognitively less demanding HL design. It corroborates the large difference in noise content of these two task designs for eliciting risk preferences.

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

[23] For more details, see Table A.3 which displays estimated regression coefficients along with calculated marginal effects.

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

		% Safe Choices	% Safe Choices on HL	% Safe Choices on OLS	% Reversals	% Reversals on HL	% Reversals on OLS	HL Switch SD
Demographic and Socioeconomic Variables	R2	0,05	0,04	0,07	0,02	0,03	0,02	0,03
All Parameters	R2	0,82	0,89	0,58	0,48	0,27	0,54	0,42
Coefficient of Risk Aversion		88,5%	88,5%	51,2%	0,0%	0,0%	0,0%	0,6%
Consistency Parameters		11,5%	11,5%	48,8%	100,0%	100,0%	100,0%	99,4%
- Stability		0,7%	1,2%	0,3%	0,3%	13,0%	0,0%	63,3%
- P(Effort)		10,7%	10,4%	48,5%	99,7%	87,0%	100,0%	36,1%

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

Cross-sectional variation in choice reversals - a strong form of choice inconsistency *within* an MPL - is explained largely by differences in the propensity to exert sufficient effort on both task designs. This is consistent with the finding that the median individual has stable risk preferences and choice inconsistency on lottery tasks is thus largely due to mistakes caused by rational inattention.[24] However, cognitive noise, captured by the standard deviation of the coefficient of risk aversion, accounts for the majority of the explained cross-sectional variation in inconsistent switching points, a more subtle form of choice inconsistency *between* MPLs. While apparent preference instability and propensity to exert sufficient effort both explain randomness in observed decisions, each has a separate role and affects the two analyzed task designs to different degrees.

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

These results illustrate the identification strategy outlined in Section 5.a.ii and complement the findings of Jagelka (forthcoming).

Another interesting result visible in Table A.3 is the lack of a relationship between the coefficient of risk aversion and choice reversals. This nuances Bruner (2017)'s claim that a negative relationship between mistakes and risk aversion is a general feature of monotone models such as the RPM.[25]

6.a.ii Determinants of Individual Choices

We next examine how well our model predicts *each individual choice*. While it is in general more difficult to predict a decision on a specific task rather than average behavior, individual choices simulated using estimated structural estimates match observed choices 85% of the time for the median individual.

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 at which an individual would be indifferent between the two lotteries. Estimating a simple linear regression, Table A.2 shows that, as implied by the model, an individual's coefficient of risk aversion being above or below the indifference threshold Θ_l^{eq} for a given choice task (henceforth referred to as the "threshold dummy") is the most significant predictor of an observed choice on that task.[26] This information alone explains 76% of the cross-sectional variation in *individual* choices on lottery tasks of the HL design. However,

[25]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, preference instability 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 risk aversion further away from the threshold level of indifference and thus a given level of preference instability 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 preference instability (let alone due to decision errors) on tasks where the threshold level of indifference varies such as the ones used in this experiment.).

[26]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.

on tasks of the OLS design it explains only 18% of the cross-sectional variation in *individual* choices on lottery tasks. Once the threshold dummy is accounted for, the inclusion of the full set of payoff-relevant task parameters (lottery payoffs and their associated probabilities) in the regression has no meaningful impact. Adding an interaction between the effort parameter and the threshold dummy does not affect the ability of our model to predict choices on HL tasks but almost triples it for OLS tasks.

The last three columns of Table A.2 show that the endogenous effort propensity (modeled as a function of relative stakes, task order, and task design) in and of itself accounts for virtually all of the explained variation in wrong choices observed in the experiment.[27] The threshold dummy and its interactions with the remaining structural parameters contribute minimally. Finally, it is noteworthy that the 18 included demographic and socioeconomic variables together predict neither observed nor wrong choices.

6.a.iii Task Design and Bias in Estimates

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

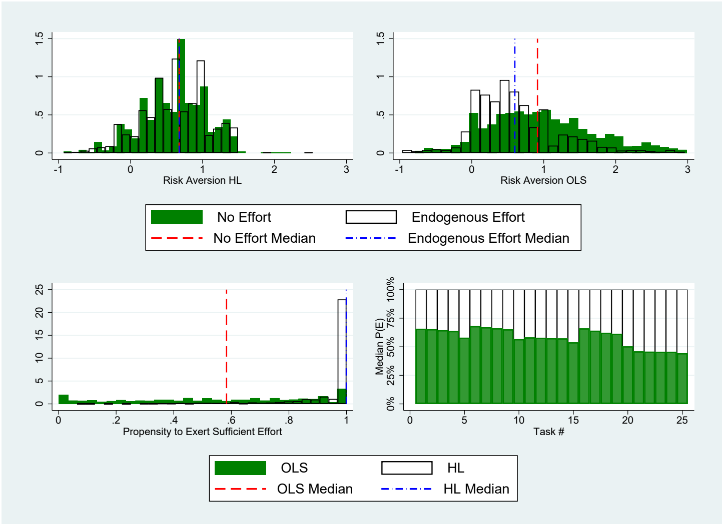
In the context of our experiment, relative stakes and fatigue only influence effort decisions on OLS tasks (i. e., their estimated average marginal effect for the median individual on HL tasks is zero). This is easily discernible from the bottom right histogram of Figure 4, which plots estimated effort propensities for the median individual on the first 25 tasks of each design. Furthermore, when we average estimated effort propensities, for each individual alternatively across the 30 tasks of the HL design and the 25 tasks of the OLS design (bottom left histogram in Figure 4), we find that most individuals exert sufficient effort *all of the time* on HL tasks whereas the median individual only exerts it approximately half of the time on OLS tasks. This suggests that there is something in the HL design which makes choosing according to latent risk preference almost effortless for the median individual.[28] Accordingly, we find that while omitting the effort decision from our model leaves the distribution of estimated risk preferences from HL choices *virtually the same* (see top left histogram of Figure 4), doing so *biases preference estimates from OLS choices for the median individual by approximately 50%* (see top right histogram of Figure 4).[29]

Andersson et al. (2016) conjecture that random decision errors will lead to an overestimation

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

[28] In contrast, the distributions of cognitive noise obtained using either task design are similar (see Figure 5). We

Figure 4: Distributions of Structural Parameters by Task Design



of risk aversion on lottery task designs in which individuals are expected to choose the riskier alternative more often than the safer one.[30] We test this hypothesis formally. For each individual, we first calculate the difference between the estimate obtained from the noisy OLS design when the effort parameter is omitted and from the HL design on which virtually all individuals choose according to their underlying risk preference. This is the bias in risk aversion resulting from a naive model which does not take into account mistakes due to inattention. We next calculate the percentage of the time that the individual would be expected to choose the riskier option on the 25 OLS-type tasks given his true (or average) risk aversion. This represents the "lopsidedness of the OLS choice tasks" for each individual. The first column of Table 2 shows that bias is indeed increasing in the lopsidedness of the lottery choice tasks towards riskier choices.

---

will discuss the implications of this finding in more detail in Section 6.b.

[29]The estimated coefficient of relative risk aversion using OLS tasks is 0.6 when endogenous effort is accounted for and 0.91 when it is excluded. On HL 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.

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

Table 2: Bias as a Function of Individual's Predicted Percentage of Risky Choices and Choice Inconsistency on OLS-Type Tasks

	Estimated Upwards Bias in CRRA Coefficient of Risk Aversion		
	(1)	(2)	(3)
Predicted % Riskier	1.92***	1.27***	1.17***
	(0.14)	(0.20)	(0.25)
P(No Effort)		1.69***	1.69***
		(0.28)	(0.30)
Predicted % Riskier * P(No Effort)		4.58***	4.78***
		(1.10)	(1.15)
Constant	0.12***	-0.13**	-0.12
	(0.036)	(0.053)	(0.065)
Observations	1,224	1,224	1,224
R-squared	0.129	0.171	0.171

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

The bias should be larger for individuals who are less likely to exert sufficient effort on the choice tasks and are thus more prone to making mistakes. In the second column we add the estimated probability of not exerting effort along with the interaction term. Both estimated coefficients are significant and positive as predicted. Bias is highest for individuals who are prone to mistakes when their true 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 0.48 increase in the bias of the estimated coefficient of relative risk aversion.[31] Given a median value of 0.68, this corresponds to an upward bias of 75%. It can be understood as the effect of design imbalance at the individual level. In addition, increasing inattention by one standard deviation is predicted to increase bias by 0.22.

Cognitive noise can also be seen as a source of random error relative to choices based on an individual's true (or average) risk preference. However, due to the symmetric nature of the shock, which is likely to cause choice inconsistency predominantly *around* an individual's average switch point (see Table 1), we do not expect to see much bias introduced by preference shocks. Nevertheless, we add the standard deviation of the coefficient of risk aversion, along with the relevant interaction term, to the regression for completeness in the third column. The estimated coefficients related to cognitive noise are statistically insignificant and the remaining estimates are

[31] The calculated marginal effect includes interaction terms calculated at mean values of the probability to exert effort.

unchanged.

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

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

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

PISA (Program for International Student Assessment) is the most important International Large-Scale Assessment (ILSA) of student achievement. The PISA exam is regularly administered to representative samples of national populations and is meant to provide a basis for international comparisons of student achievements. PISA assesses three core domains; mathematics, reading and science, but the mathematics component raises a particularly high level of interest as virtually all OECD countries would like to stimulate student enrollment in mathematical and scientific academic subjects (STEM).

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

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

To answer the first question, we regress PISA numeracy scores and high school GPA on the individual specific effort propensity, controlling for other skills, preferences, and characteristics. To answer the second question, we regress PISA numeracy scores and high school GPA on self-reported high school engagement.[32] To facilitate comparison, we standardize all variables apart

[32]High school engagement is calculated based on self-reports (hours spent on homework, handing in homework on

from sex. The results are summarized in Table 3.

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

These results are interesting for many reasons. First, they show that our estimate of low-stakes effort propensity has external validity. This suggests that we may be capturing a more general behavioral tendency, *low-stakes motivation*. The estimated marginal effect of low-stakes motivation on PISA test scores is meaningful in magnitude. Increasing effort by one standard deviation, holdings self-reported skills, personality, and sex constant, is predicted to add an additional 12 points on the 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), 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).[33] Second, they indicate a fundamental disconnect between effort provided in low stakes environments and effort exerted in situations with potentially large impacts on a person's future.[34]

time, self-reported effort in high school).

[33]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.

[34]The fact that low-stakes motivation retains some predictive power for high school GPA may be an artifact of the

Table 3: Predictive Power of Low-Stakes and High-Stakes Motivation on the PISA Achievement Test and High School GPA

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

Standard errors in parentheses.

*** p<0.01, ** p<0.05

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

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

latter being a sum of many constituent task performances, some of which can be perceived as low-stakes. While our estimated low-stakes motivation (effort propensity) is a statistically significant predictor of high-stakes motivation (high school engagement), the correlation between the two is very low (<0.07).

6.b Stability of Individuals' Preferences

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

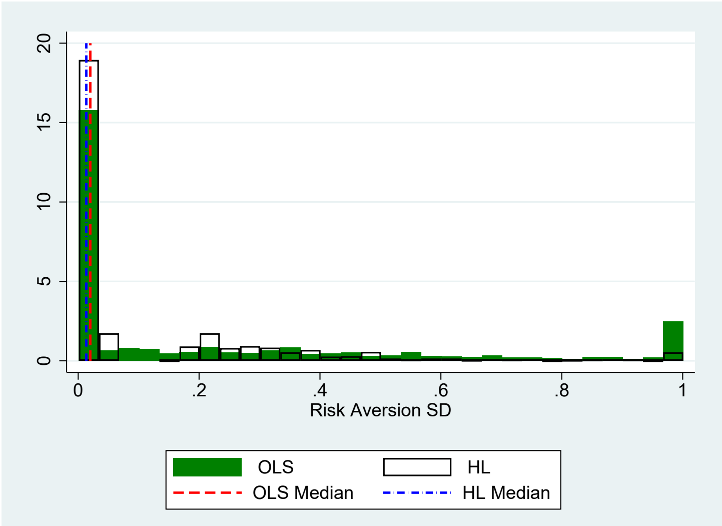
The concept of unstable preferences is not standard in the economic literature and indeed there is a limit to how much preferences can plausibly fluctuate within a short time interval. Using the same dataset but estimating a model *without* endogenous effort, Jagelka (forthcoming) shows that "for the median individual [by estimated risk aversion], choice inconsistency generated by the estimated preference shocks is concentrated within one or two cells from the switch point implied by constant preferences set at their average value". However, existing estimates of the scale of the error shocks are still large in absolute terms.[35]

One of the contributions of this paper is to show that after accounting for differences in situations, preferences become stable for the median individual. A particular task design is a situation. Preference instability estimated using only tasks of the same design is low. Furthermore, preference instability estimated separately on HL and on OLS-type tasks is similar while the likelihood of exerting sufficient effort is different (see Figure 5 and the bottom left histogram of Figure 4, respectively). The fact that preference instability is the same across task designs while mistakes due to inattention vary suggests that the stability of preferences is an individual characteristic while decision errors are due to rational inattention, responsive to incentives.

Once the decision to exert effort is incorporated into the model, the median individual has stable risk preferences even when all 55 available lottery choice tasks are used for estimation. Com-

[35] To see this consider that for normally distributed shocks with a standard deviation of 0.4 (approximately the midpoint of previous estimates, see Apesteguia and Ballester, 2018; Apesteguia, Ballester, and Gutierrez, 2020; Jagelka, forthcoming), an individual with an estimated coefficient of relative risk aversion of 0.7 would be predicted to alternatively behave as *risk-seeking* or *highly risk-averse* more than 5% of the time.

Figure 5: Distributions of Estimated Cognitive Noise Parameter by Task Design



bined 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 on the OLS design. This suggests that the failure to account for differences in situations results in a misattribution of some choice inconsistency to cognitive noise.

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

6.c Welfare Impact

One of the advantages of using a structural model is that it allows us to quantify the cost of choices which are inconsistent with an individual's true (or average) preference due to rational inattention or cognitive noise. In the context of our experiment, we calculate each individual's "ideal" choice based on estimated true (or average) risk aversion.[36] The welfare loss on each choice task is calculated as the difference in certainty equivalents between an individual's actual and "ideal" choice given his estimated coefficient of risk aversion Θ_i.

The average individual loses approximately 2% of utility or $1 per task due to choices inconsistent with his true risk preference.[37] This loss is almost exclusively due to choices on OLS-type tasks. We interpret it as the additional cost of effort, per task, imposed by the OLS design as opposed to the HL design.[38] This is smaller than the welfare loss attributed by Choi et al. (2014) to low quality decision-making. However, it is important to remember that choice tasks in this experiment were designed to be clear and simple in order to elicit true preferences. The fact that even in such a simple controlled laboratory setting inattention and cognitive noise are associated with lost welfare, suggests that welfare loss is magnified in real-world settings where the level of complexity is much higher.

Table A.9 shows that 60% of the cross-sectional variation in estimated welfare loss is explained by individual heterogeneity in effort and cognitive noise. For comparison, nearly two dozen demographic and socioeconomic variables explain less than 5% of the cross-sectional variation. A one standard deviation increase in the likelihood of not exerting sufficient effort translates to a loss of 44 cents per task whereas a one standard deviation increase in cognitive noise results in a loss of 23 cents per task. Almost 75% of explained welfare loss is attributable to mistakes resulting from inattention. This is consistent with findings presented in the previous sections: endogenous effort is a larger source of choice inconsistency than cognitive noise.

Finally, we construct a **Consistency Index** which reflects an individual's overall propensity to make inconsistent choices. It combines the estimated consistency parameters in our model—the standard deviation of the coefficient of risk aversion and an individual's average estimated propensity to exert sufficient effort.[39] The Consistency Index provides a convenient characterization of the degree of inconsistency of each individual's choices whether it is due to apparent preference instability (imperfect self-knowledge) or voluntary inattention. Among the 1,224 individuals in our sample, the Consistency Index ranges from 11 to 100, with a mean and median

[36] Choice mistakes due to inattention quantify the cost of inattention. Cognitive noise results in lost welfare so long as it is an artifact of an individual's imperfect self-knowledge which makes him guess around the true preference. Both sources of choice inconsistency may nevertheless reflect rational behavior if there are costs to effort and introspection.

[37] Losses are 0.2% of utility at the 5th percentile and 4.8% at the 95th percentile. The per task loss is the relevant object for this experiment as each individual was compensated for one of the choice tasks, drawn at random.

[38] Recall that preferences are estimated to be stable for the median individual.

[39] All parameters are first adjusted to a scale of 0-100 with 100 implying the highest choice consistency. The Consistency Index is an average of these values.

just below 80. Figure A.3 plots its distribution. Table A.10 shows that the Consistency Index is a strong predictor of welfare loss on the experimental tasks.

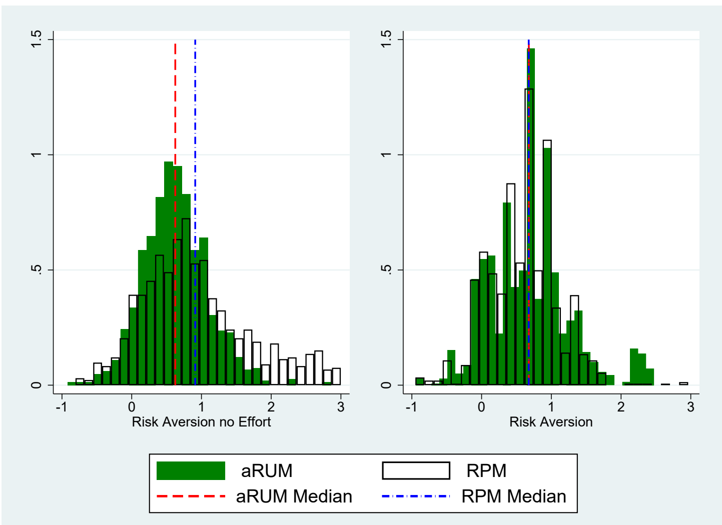
6.d Comparison with the Random Utility Model

Previous structural research on estimating risk preferences overwhelmingly relied on variants of the aRUM model. Due to the non-monotonicities of the aRUM pointed out by Apesteguia and Ballester (2018), we view the RPM as a theoretically more sound alternative. Nevertheless, given the existence of a large body of research using the aRUM, we consider it worthwhile to compare estimates using the two competing error specifications embedded within our endogenous effort framework.

Jagelka (forthcoming) 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.

We corroborate these results when we estimate risk aversion *without taking into account the initial effort decision*. In this case, the entire distribution of the estimated coefficient of relative risk aversion is skewed to the right when using preferences shocks rather than additive utility shocks (see the left histogram of Figure 6 below).[40]

Figure 6: Distributions of Structural Parameters Estimated Using All Tasks with Alternatively the RPM and aRUM Error Structure



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

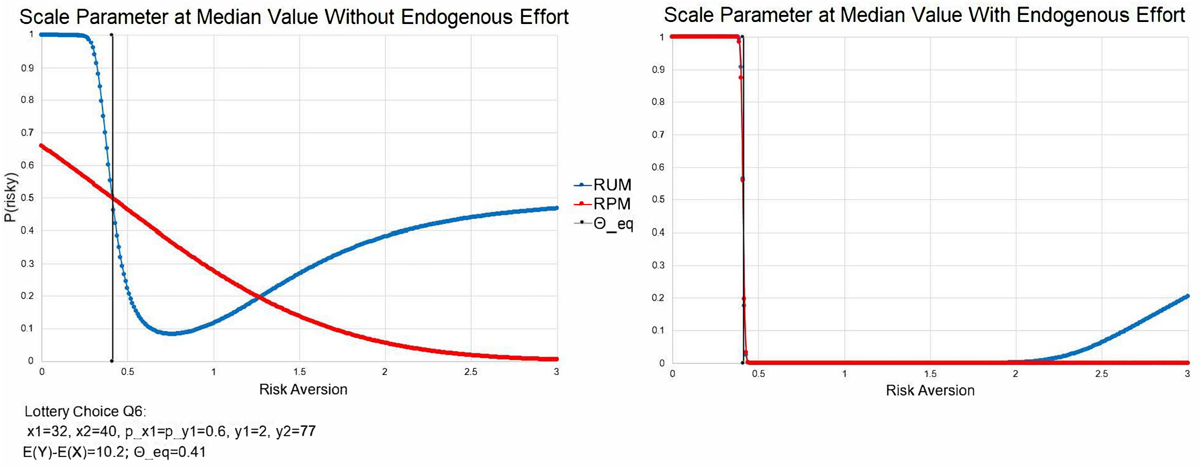
Once we estimate our model with endogenous effort, the non-monotonicity of the aRUM becomes empirically irrelevant, at least in the context of our experimental sample. The distributions of the coefficient of relative risk aversion estimated using either preference shocks or additive utility shocks converge (see the right histogram of Figure 6). 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. After taking into account endogenous effort, one could thus put risk preference estimation within an aRUM framework in the same category as time preference estimation: theoretically problematic but empirically largely irrelevant.[41]

To illustrate this point, we take as an example the 6th choice task of the HL design contained in our data. In Figure 7 we plot the predicted probability of choosing the riskier lottery Y under RPM and under aRUM for values of risk aversion between 0 and 3 when the variance of the scale parameter σ_i is set at the median estimate using alternatively a model *without* endogenous effort (left) and our full model *with* endogenous effort (right). In either case, both the RPM and aRUM curves are initially decreasing, in line with the intuition that a more risk averse individual should be predicted to choose the riskier option with a lower probability. The curves cross at the threshold level of indifference for this choice task ($\Theta_l^{eq} = 0.41$) where by definition the expected utilities of the two lotteries are equal and both models correctly predict that the probability of choosing either option is 0.5. The graph on the left assumes error shocks of a magnitude estimated for the median individual *when the effort decision is omitted*. The RPM curve continues to decrease monotonically while the aRUM curve reverts with risk aversion still below one (and thus while still at moderate and empirically frequent values of Θ_i). It resembles Figure 1 in Apesteguia and Ballester (2018), which they use to illustrate the non-monotonicity problem of the aRUM. The graph on the right assumes error shocks of a magnitude estimated for the median individual *when the effort decision is endogenized*. Conditional on effort, the probability of choosing the riskier option becomes almost degenerate (deterministic). While it increases again for the aRUM, it does so at a much higher value of risk aversion. The non-monotonicity problem becomes practically irrelevant in terms of the empirical estimation of risk aversion using our data: the estimated distributions of the coefficient of risk aversion converge

[41] 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.

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.

Figure 7: The RPM vs. aRUM Likelihood of Selecting the Riskier Lottery on the 6th Lottery Choice Task Assuming CRRA Utility



6.d.i Mean-Variance Utility

As noted in the recent literature, it is possible to obtain a monotonic version of the Random Utility model if one is willing to assume a form of mean-variance (MV) utility function, such as is commonly done in the finance literature, or if one uses a Taylor-series approximation of a concave utility function with negligible third-order derivative (such as in Cohen and Einav, 2007; Barseghyan et al., 2013). This approach delivers a parameter that is not directly interpretable as a coefficient of relative risk aversion. It is nevertheless possible to evaluate how an individual's sample rank by implied risk (or variance) aversion under MV utility compares to his rank obtained from aRUM and RPM estimates with CRRA utility.

In our context, MV utility yields the following decision equations which individual i uses to compare the safer lottery X and the riskier lottery Y:

$$MV_i(X) = mean(X) - \gamma_i \cdot var(X) + \varepsilon_{i,X}^{MV} \tag{19}$$

and

$$MV_i(Y) = mean(Y) - \gamma_i \cdot var(Y) + \varepsilon_{i,Y}^{MV} \tag{20}$$

where $mean(X)$ is the expected value of lottery X, $var(X)$ is the variance of lottery X, γ_i represents risk aversion of indiviual i, and $\varepsilon_{i,X}$ and $\varepsilon_{i,Y}$ are individual i's draws of the error term for lottery X and Y respectively. Assuming the two shocks are independent and normally distributed random variables, the probability that individual i prefers the riskier lottery Y on choice task l is:

$$\begin{aligned} P(YP_{i,l}=1)^{MV} &= P[MV_i(Y) > MV_i(X)] \\ &= P\left[\varepsilon_{i,Y}^{MV} - \varepsilon_{i,X}^{MV} > MV_i(X) - MV_i(Y)\right] \\ &= \Phi\left[\frac{MV_i(Y) - MV_i(X)}{\sigma_i^{MV}}\right] \end{aligned} \tag{21}$$

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

While the risk aversion parameter obtained using mean-variance utility is not directly comparable to the CRRA coefficients, its rank correlation with both the CRRA, RPM and aRUM estimates is very high (0.93 and 0.95 respectively), see Table 4.

Table 4: Rank Correlations of Estimates of Risk Aversion Obtained Using the RP-CRRA, RU-CRRA, and RU-Mean-Variance Models

	RPM: CRRA	aRUM: CRRA	aRUM: MV
RPM: CRRA	1.00		
aRUM: CRRA	0.97	1.00	
aRUM: MV	0.93	0.95	1.00

7 Concluding Discussion

We develop a general framework that accounts for endogenous effort and cognitive noise which confound estimates of preferences, skills, and other latent personal attributes (PSAs) based on observed behavior. We exploit variation in the design of choice tasks for eliciting risk preferences to demonstrate how our model can be used to detect noise in observed choices and to de-bias preference estimates.

Our model's defining feature is that it formally incorporates an individuals' initial decision about whether to spend effort on a given task, based on that task's readily and effortlessly available characteristics. If the individual decides to exert effort, he chooses the expected utility maximizing option, taking into account an error shock reflecting cognitive noise. If not, he acts according to a decision heuristic.

We estimate the model using experimental data from a representative sample of over 1,200 individuals, each of whom made 55 binary choices on incentivized tasks of two designs often used for eliciting risk preferences. Estimated structural parameters explain observed choices well. Their explanatory power dwarfs that of a wide range of demographic and socioeconomic variables.

The availability of a long panel allows us to study preferences and decision noise at the individual level. One of the advantages of having individual-specific estimates is that these may be used to test the external validity of the structural parameters of a model. We find that estimated effort propensity is particularly predictive of an individual's performance in other low-stakes environments, even when controlling for other PSAs and demographics. This suggests that it captures a more general individual characteristic, low-stakes motivation. We show that a one standard deviation increase in low-stakes motivation could affect the PISA numeracy ranking of a mid-performing country by approximately 9 places (a 40% jump in the rankings).

In addition, we provide evidence that the propensity to exert effort in *low-stakes settings* is fundamentally different from the propensity to exert effort in *high-stakes settings*. To this end, we construct a measure of high-stakes motivation based on self-reported school engagement in high school. While estimated low-stakes motivation (effort propensity) is a statistically significant predictor of high-stakes motivation (high school engagement), the correlation between the two is very low (<0.07). Furthermore, we find that school engagement is not a statistically significant predictor of performance on PISA numeracy scores after controlling for self-reported skills. However, it is an excellent predictor of a high-stakes outcome: high school GPA.

We show that both endogenous effort and cognitive noise drive inconsistency in observed choices and fulfill specific roles. Rational inattention is the main source of noise on the examined tasks. Individuals are less likely to exert the effort necessary to correctly evaluate the choice at hand when mental processing costs and fatigue are high and when stakes associated with making a correct choice are low. Once the first-stage decision to exert effort is incorporated, the underestimation of risk aversion for the mean individual when using an aRUM documented by Apesteguia and Ballester (2018) disappears. At least in the context of this experiment, proper estimation of the first stage is empirically more important than the placement of the error term. We nevertheless view the RPM as the proper framework to use in preference estimation due to its superior theoretical properties and to the intuitive interpretation of apparent preference instability as a manifestation of imperfect self-knowledge.

We find that cognitive noise is invariant to task design, unlike mistakes due to inattention. This suggests that the stability of preferences is an individual characteristic while decision errors are driven by rational inattention, which is responsive to the net benefits of exerting effort determined by the characteristics of a choice situation. Furthermore, estimated preferences are stable for the median individual once we account for endogenous effort. This is good news for traditional economic theory.

Nevertheless, a non-negligible part of the population behaves as if they had not only one true value of risk preference but rather a distribution centered around it, consistent with the basic assumption underlying the Random Preference Model. Such preference instability may be attributed to cognitive noise due to imperfect self-knowledge.

Empiricists currently use a plethora of elicitation instruments for PSAs, featuring a number of design variations, with no 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*—HL and binarized OLS—for eliciting risk preferences which were previously often used interchangeably. On the one hand, we show that choices on tasks of the more intuitive HL design largely reflect an individual's latent risk preference. 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 reveals that half of the explained cross-sectional variation in average choices on the more complex OLS tasks can be attributed to random decision-making due to insufficient effort (in which case choices are uninformative about latent risk preference), implying a signal-to-noise ratio of approximately 1/2. Omitting the initial effort decision results in estimates of risk aversion biased by 50% for the median individual on OLS-type choice tasks. The bias is higher for individuals who have a high propensity to make mistakes and whose true risk preference would disproportionately make them choose the riskier alternative.

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

Does this mean that HL-type tasks are better suited than OLS tasks to elicit risk preferences and should thus be used exclusively? Not necessarily. In the context of this experiment the two types of choice tasks are complementary and can be used in conjunction to extract more information assuming an appropriate econometric framework is used. The calculated indifference thresholds displayed in Tables A.6 and A.7 illustrate that while the HL design covers the most common levels of risk-aversion, information from OLS tasks can be used to narrow down the interval within which an individual's true coefficient of risk aversion lies and to capture more extreme behavior at the high end of the distribution.[42] However, OLS-type tasks will only provide valid estimates if choice inconsistency is properly accounted for. The HL design augmented to cover a wider range of risk preferences would seem recommendable, especially if reduced-form techniques are to be relied upon in estimation.

The obvious question is: What causes the large difference in individuals' effort decisions on the two available task designs? The HL and OLS lottery choices differ in multiple aspects. Ascertaining their individual contributions to the perceived costs of exerting sufficient effort is beyond the scope of this paper as it would require random variation in each examined design feature. However, as discussed in Section 4, the *ensemble* of features of the HL design work to minimize the per-task effort required to choose according to one's latent risk preference: on the first task of each HL-type Multiple Price List (MPL), the safer lottery offers a higher expected value than the riskier one, making for a rather simple choice for most individuals. On subsequent tasks within the MPL, it is obvious that the probability of receiving the higher payment in each lottery increases. As the riskier option becomes progressively more attractive, an individual can think about, and clarify, his own risk preference. In a sense, he is being gently nudged towards the point at which he wishes to switch. This makes for a very simple setting to elicit preferences, with low mental processing costs per choice and low cognitive demand. The amount of effort required to choose according to latent preferences on a given task, *sufficient effort* according to our definition, is thus likely very low.

On tasks of the OLS design, the pattern between subsequent tasks of a given MPL is not easily discernible. This makes the choices less intuitive and potentially requiring thought and varying amounts of effort, depending on one's ease of processing the tasks which in turn likely depends

[42] This is a feature of the particular parametrization of the HL tasks used in this experiment (which, however, is very standard in the literature) than of the design itself. The choices could be modified so that HL tasks cover a wider range of risk aversion and have overlapping intervals of thresholds of indifference. This would plausibly keep the advantage of the HL tasks in terms of reflecting true preferences and eliminate their limitations as seen in the context of this experiment.

on cognitive and non-cognitive skills. This is precisely the context in which one can expect differentiation in the amount of mistakes made based on unobserved heterogeneity. It is reflected in the wide dispersion of our estimated effort propensities on OLS 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 PSA 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 true 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. We find that a Consistency Index condensing each individual's estimated consistency parameters into a single indicator is highly predictive of welfare loss in this simple experimental setting, with the average individual leaving 2% of welfare on the table. Losses in real-world decisions are likely magnified.

It has been documented that ignoring the stochastic components of decisions can lead to biased inference on preferences in a variety of fields such as education or healthcare. For example Hastings, Kane, and Staiger (2006) initially inferred that low-income families placed little weight on academic quality when choosing schools for their kids but Hastings and Weinstein (2008) show that the observed choices reflected noise rather than true preferences. Low-income parents faced a higher cost of acquiring the information necessary to act on their preferences just as low ability individuals seem to face a higher cost of evaluating the options in some of the tasks in this experiment (see Jagelka, forthcoming, on the link between mistakes and cognitive ability). In both cases, one needs to account for instances in which a person's preferred choices differ from his actual ones in order to infer true underlying preferences.

These findings open up multiple avenues for future research. We show that observed individual decisions reflect a mix of signal and noise to different degrees on seemingly similar choice tasks. Future applications of our model should aim to disentangle the impact of particular task design features on the noise content of observed choices. This will require collecting data with random variation in the design features of interest. New datasets should also collect decision times and use them, in conjunction with endogenous effort estimates, to clarify the controversial relationship between effort provision and survey time. In addition, it is desirable to compare structural effort and cognitive noise estimates to reduced-form ways of detecting low quality responses such as asking individuals to self-report the overall reliability of their answers or asking them how sure they are of a given choice.

The ultimate question is the external validity of estimated preferences, effort, and cognitive noise. The importance of low-stakes and high-stakes motivation in real-world settings merits further study. The predictive power of preference estimates on inequality in outcomes should be re-evaluated once decision noise is accounted for. Existing mixed results regarding the external validity of estimates elicited in laboratory settings may be an artifact of varying amounts of noise in observed choices in different contexts rather than of context-dependent preferences. If this is the case, the way forward lies in better understanding where the noise comes from. 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.

8 Bibliography

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.

Almlund, M., A.L. Duckworth, J.J. Heckman, and T.D. Kautz. 2011. *Personality psychology and economics* [in en]. National Bureau of Economic Research.

Andersen, S., G.W. Harrison, M.I. Lau, and E.E. Rutström. 2008. “Eliciting risk and time preferences.” *Econometrica* 76 (3): 583–618.

Andersson, O., H.J. Holm, J.R. Tyran, and E. Wengström. 2016. “Risk aversion relates to cognitive ability: preferences or noise?” *Journal of the European Economic Association* 14 (5): 1129–1154.

Andersson, O., H.J. Holm, J.R. Tyran, and E. Wengström. 2020. “Robust inference in risk elicitation tasks” [in en]. *Journal of Risk and Uncertainty* 61:195–209.

Apesteguia, J., and M.A. Ballester. 2018. “Monotone stochastic choice models: The case of risk and time preferences” [in en]. *Journal of Political Economy* 126 (1): 74–106.

Apesteguia, J., M.A. Ballester, and A. Gutierrez. 2020. *Random models for the joint treatment of risk and time preferences* [in en]. Working Paper.

Barseghyan, Levon, Francesca Molinari, Ted O’Donoghue, and Joshua C Teitelbaum. 2013. “The nature of risk preferences: Evidence from insurance choices.” *American economic review* 103 (6): 2499–2529.

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.

Belzil, C., and M. Sidibé. 2016. "Internal and External Validity of Experimental Risk and Time Preferences." *IZA Discussion Paper* 10348.

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.

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.

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.

Coffman, Katherine B., and David Klinowski. 2020. "The impact of penalties for wrong answers on the gender gap in test scores." *Proceedings of the National Academy of Sciences* 117 (16): 8794–8803.

Cohen, Alma, and Liran Einav. 2007. "Estimating risk preferences from deductible choice." *American economic review* 97 (3): 745–788.

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.

Deming, David J. 2017. "The growing importance of social skills in the labor market." *The Quarterly Journal of Economics* 132 (4): 1593–1640.

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. Forthcoming. "Accounting for Individual-Specific Reliability of Self-Assessed Measures of Economic Preferences and Personality Traits." *Journal of Political Economy: Microeconomics.*

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.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. "Global evidence on economic preferences." *The Quarterly Journal of Economics* 133 (4): 1645–1692.

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

Gabaix, Xavier. 2019. "Behavioral inattention." In *Handbook of behavioral economics: Applications and foundations 1,* 2:261–343. Elsevier.

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.

Guadalupe, Maria. 2007. "Product Market Competition, Returns to Skill, and Wage Inequality." *Journal of Labor Economics* 25 (3): 439–474. https://doi.org/10.1086/513299.

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. Elisabet Rutström. 2008. "Risk aversion in the laboratory." In *Risk aversion in experiments,* 41–196. Emerald Group Publishing Limited.

Hastings, Justine S, Thomas J Kane, and Douglas O Staiger. 2006. "Preferences and Heterogeneous Treatment Effects in a Public School Choice Lottery." Working Paper, Working Paper Series 12145. National Bureau of Economic Research. https://doi.org/10.3386/w12145.

Hastings, Justine S, and Jeffrey M Weinstein. 2008. "Information, school choice, and academic achievement: Evidence from two experiments." *The Quarterly journal of economics* 123 (4): 1373–1414.

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.

Heckman, J.J., J. Stixrud, and S. Urzua. 2006. "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior." *Journal of Labor economics* 24 (3): 411–482.

Heidhues, Paul, and Botond Kőszegi. 2017. "Naïveté-Based Discrimination*." *The Quarterly Journal of Economics* 132 (2): 1019–1054. https://doi.org/10.1093/qje/qjw042.

Hey, J.D., and C. Orme. 1994. "Investigating generalizations of expected utility theory using experimental data." *Econometrica: Journal of the Econometric Society,* 1291–1326.

Holt, C.A., and S.K. Laury. 2002. "Risk aversion and incentive effects." *American economic review* 92 (5): 1644–1655.

Jagelka, Tomáš. Forthcoming. "Are economists' preferences psychologists' personality traits? A structural approach." *Journal of Political Economy.*

Kahneman, D. 2011. *Thinking, fast and slow.* macmillan.

Khaw, Mel W., Ziang Li, and Michael Woodford. 2021. "Cognitive imprecision and small-stakes risk aversion." *The review of economic studies* 88 (4): 1979–2013.

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.

Lang, Frieder R., Dennis John, Oliver Lüdtke, Jürgen Schupp, and Gert G. Wagner. 2011. "Short assessment of the Big Five: robust across survey methods except telephone interviewing." *Behavior research methods* 43 (2): 548–567. https://doi.org/10.3758/s13428-011-0066-z.

Loomes, G., and R. Sugden. 1995. "Incorporating a stochastic element into decision theories." *European Economic Review* 39 (3-4): 641–648.

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

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

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

Miller, L., D.E. Meyer, and J.T. Lanzetta. 1969. "Choice among equal expected value alternatives: Sequential effects of winning probability level on risk preferences." *Journal of Experimental Psychology* 79 (3p1): 419.

Soto, C.J., and O.P. John. 2017. "The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power." *Journal of personality and social psychology* 113 (1): 117.

Steverson, Kai, Adam Brandenburger, and Paul Glimcher. 2019. "Choice-theoretic foundations of the divisive normalization model." *Journal of Economic Behavior & Organization* 164:148–165. https://doi.org/https://doi.org/10.1016/j.jebo.2019.05.026.

Thurstone, L.L. 1927. “A law of comparative judgment.” *Psychological review* 34 (4): 273.

Todd, Petra E, and Weilong Zhang. 2020. “A dynamic model of personality, schooling, and occupational choice.” *Quantitative Economics* 11 (1): 231–275.

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 Sample Descriptive Statistics

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%

A.b Choice Tasks

Figure A.1: Lottery Choice Tasks - MPL Design

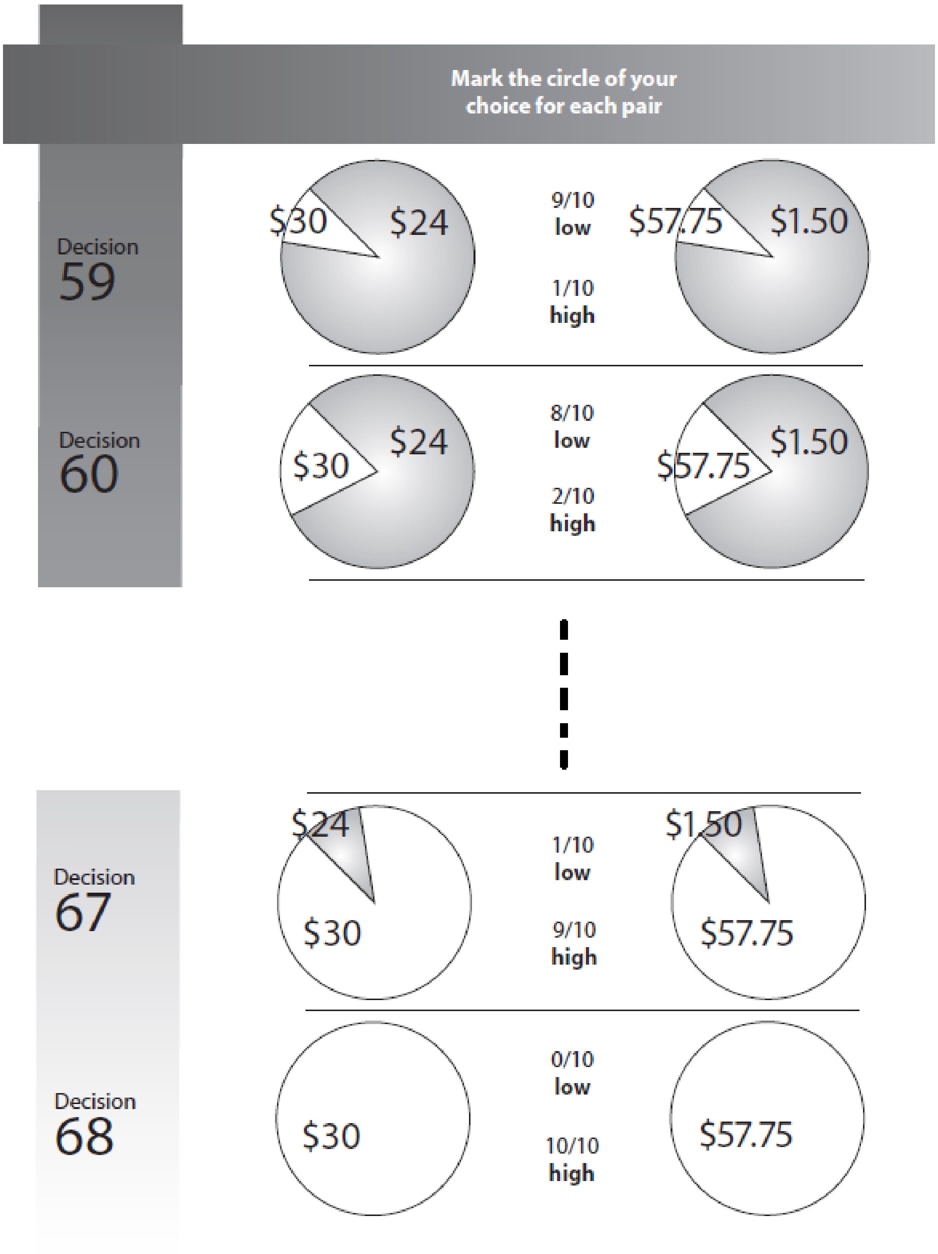


Figure A.2: Lottery Choice Tasks - OLS design

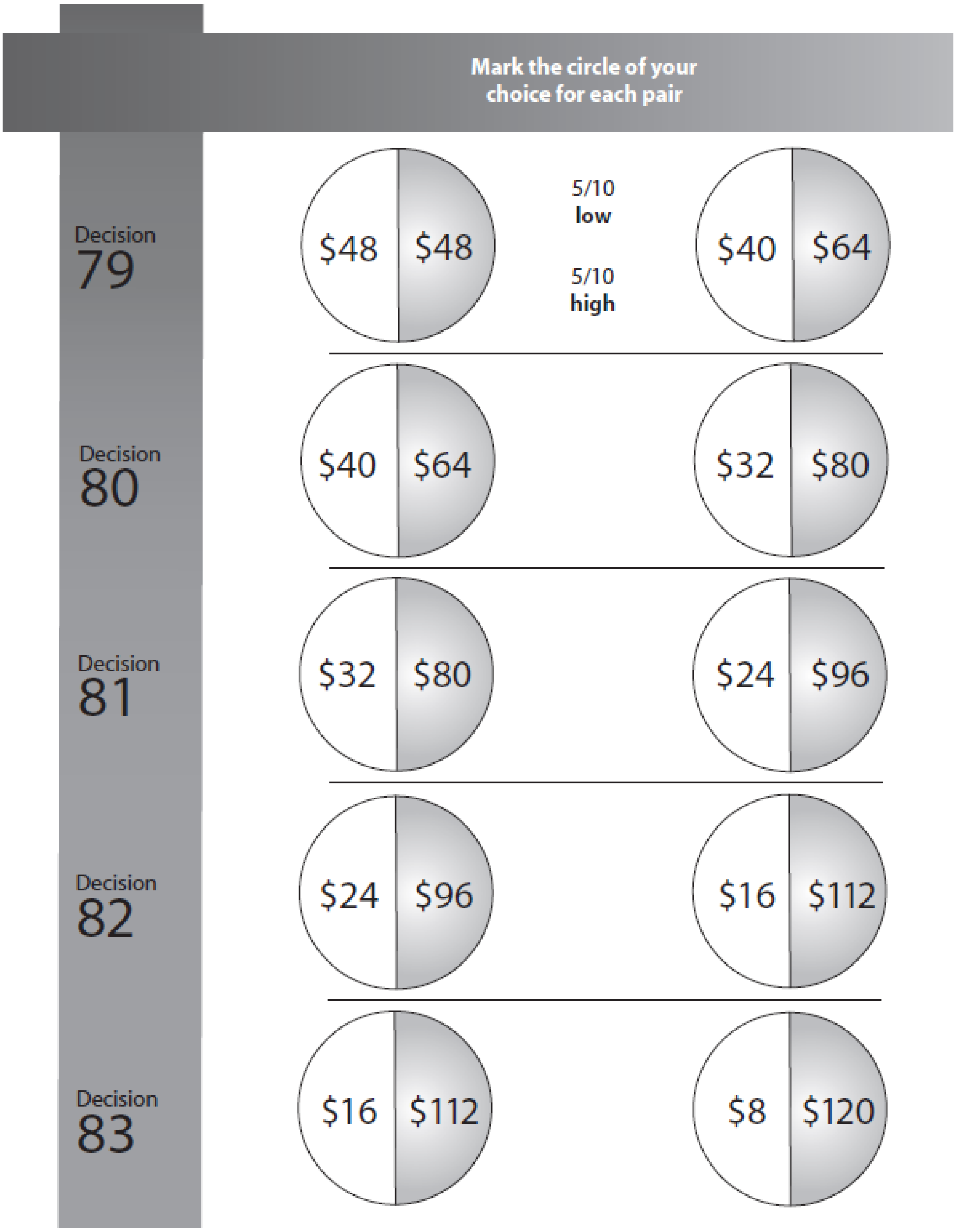


Table A.2: Explanatory Power of Individual Determinants of Lottery Choices

		Observed Choices			Wrong Choices		
		All	HL	OLS	All	HL	OLS
Demographic and Socioeconomic Variables	R2	0.00	0.00	0.01	0.00	0.00	0.00
Threshold Dummy	R2	0.44	0.76	0.18	0.01	0.00	0.00
P(Effort)	R2	0.00	0.00	0.00	0.36	0.22	0.32
P(Effort) * Threshold Dummy	R2	0.64	0.81	0.44	0.37	0.24	0.32
Full Set of Regressors	R2	0.66	0.83	0.46	0.39	0.29	0.34

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

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

	% Safe Choices	% Safe Choices on HL	% Safe Choices on OLS	% Reversals	% Reversals on HL	% Reversals on OLS	HL Switch SD
Risk Aversion	43***	35.2***	54.8***	-0.4	-0.2**	-1.1*	0
	(0.82)	(0.41)	(1.70)	(0.33)	(0.11)	(0.60)	(0.04)
Risk Aversion SD	6.7***	0.7	14.6***	3.9***	2***	4.8***	2.8***
	(1.94)	(1.52)	(4.27)	(0.78)	(0.42)	(1.50)	(0.15)
Risk Aversion * SD	-13.2***	-8.7***	-17.1***	0.6	-0.3	2.5	-0.8***
	(1.92)	(1.53)	(4.27)	(0.77)	(0.42)	(1.50)	(0.16)
P(No Effort)	44.7***	40.2***	55***	17.9***	2.7***	19.1***	2.2***
	(1.03)	(0.76)	(1.02)	(0.42)	(0.21)	(0.36)	(0.08)
Risk Aversion * P(No Effort)	-59.9***	-68.5***	-67.7***	0.8	9.6***	-2.2**	0.7***
	(1.38)	(1.33)	(1.45)	(0.56)	(0.37)	(0.51)	(0.14)
Effect of Increasing Each Structural Parameter by One Standard Deviation							
- Risk Aversion	13.1	14.7	10.4	-0.1	0.1	-0.9	0.0
- Risk Aversion SD	-0.2	-0.7	0.7	0.7	0.3	1.0	0.4
- P(No Effort)	1.4	0.0	2.2	2.6	1.2	2.5	0.4

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: The rows display the coefficient of the regression of the moment listed in each column title on the full set of structural parameter estimates, including interactions. Standard errors are in parentheses. The probability of exerting effort, P(Effort), is averaged over the tasks of the relevant design (all, HL, OLS) for each individual. P(No Effort)=1−P(Effort). The effect of increasing each structural parameter by one standard deviation takes into account relevant interaction terms calculated at the average values of the variables each parameter is interacted with. The analysis excludes individuals with an estimated coefficient of risk aversion of below −2 and above +2. This leaves 1,124 individuals or over 90% of the sample.

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

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

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

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

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

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

Standard errors in parentheses

*** p<0.01, ** p<0.05

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

Table A.6: Indifference Thresholds and Observed Sample Proportions of Risky Choices on MPL Type Choice Tasks

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

Table A.7: Indifference Thresholds and Observed Sample Proportions of Risky Choices on OLS-Type Choice Tasks

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

A.c Effort by Task on the HL and OLS Design

Table A.8: Predicted Probability to Exert Sufficient Effort at the 25th, 50th, and 75th Percentile by Choice Task

	HL			OLS		
	P(Effort) by Percentile			P(Effort) by Percentile		
Question	25%	50%	75%	25%	50%	75%
1	1.00	1.00	1.00	0.29	0.66	0.98
2	1.00	1.00	1.00	0.28	0.65	0.98
3	0.99	1.00	1.00	0.27	0.65	0.98
4	0.97	1.00	1.00	0.26	0.64	0.98
5	0.96	1.00	1.00	0.18	0.58	0.97
6	0.98	1.00	1.00	0.34	0.68	0.97
7	1.00	1.00	1.00	0.33	0.67	0.97
8	1.00	1.00	1.00	0.32	0.66	0.96
9	1.00	1.00	1.00	0.31	0.65	0.96
10	1.00	1.00	1.00	0.18	0.57	0.92
11	1.00	1.00	1.00	0.21	0.58	0.94
12	1.00	1.00	1.00	0.21	0.58	0.94
13	1.00	1.00	1.00	0.21	0.58	0.94
14	0.98	1.00	1.00	0.21	0.57	0.94
15	0.97	1.00	1.00	0.17	0.54	0.94
16	0.99	1.00	1.00	0.29	0.66	0.96
17	1.00	1.00	1.00	0.26	0.64	0.96
18	1.00	1.00	1.00	0.25	0.62	0.95
19	1.00	1.00	1.00	0.24	0.61	0.94
20	1.00	1.00	1.00	0.15	0.50	0.91
21	1.00	1.00	1.00	0.11	0.46	0.85
22	1.00	1.00	1.00	0.11	0.46	0.85
23	1.00	1.00	1.00	0.11	0.46	0.85
24	1.00	1.00	1.00	0.11	0.46	0.85
25	1.00	1.00	1.00	0.08	0.44	0.83
26	1.00	1.00	1.00			
27	1.00	1.00	1.00			
28	1.00	1.00	1.00			
29	1.00	1.00	1.00			
30	1.00	1.00	1.00			

Table A.9: Determinants of Welfare Loss

		Welfare Loss ($)
Demographic and Socioeconomic Variables	R2	0.03
All Parameters	R2	0.61
Risk Aversion	%	18.5 %
	Marginal Effect	$ 0.12
Risk Aversion SD	%	6.9 %
	Marginal Effect	$ 0.23
P(No Effort)	%	74.5 %
	Marginal Effect	$ 0.44

Notes: The rows labeled "R2" list the R2 of the regression of each individual's calculated welfare loss alternatively on 18 demographic and socioeconomic variables or on the 3 estimated structural preference and consistency parameters, including interactions. Demographic variables include the students' 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. The rows "%" represent the relative explanatory power of each individual parameter, expressed as a percentage of total explained variation. The rows "Marginal Effect" represent the impact of increasing each individual parameter by one standard deviation on welfare loss. It takes into account relevant interaction terms calculated at the average values of the variables each parameter is interacted with.

Figure A.3: Population Distribution of the Consistency Index

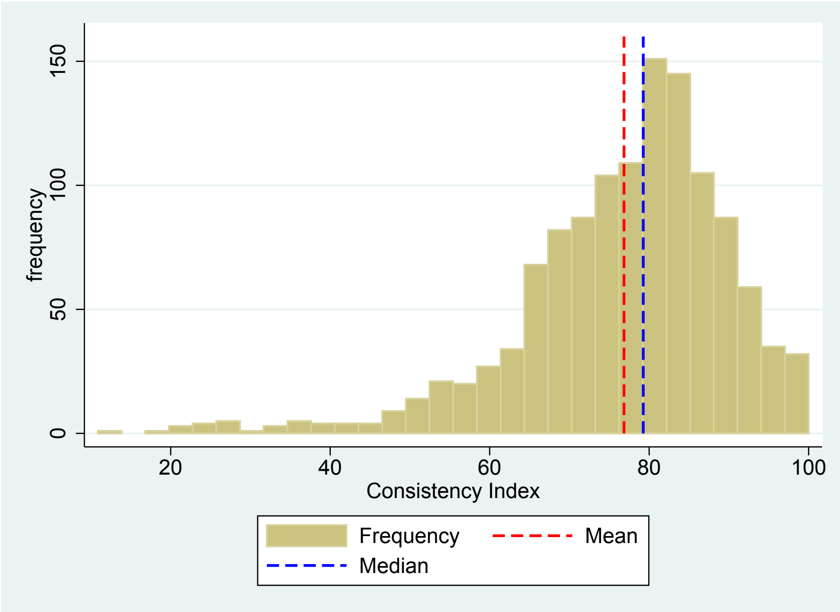


Table A.10: Impact of the Consistency Index on Estimated Welfare Loss

	Observations	Welfare Loss	
		$	%
Consistency Index Coefficient	920	-0.04	-0.09
		(0.08)	(0.00)
Consistency Index R2		0.38	0.50
Consistency Index			
<50	43	$ 2.4	5.23 %
50 - 60	117	$ 1.55	3.80 %
60 - 70	122	$ 1.12	2.73 %
70 - 80	226	$ 0.81	1.88 %
80 - 90	303	$ 0.44	1.04 %
90 - 100	109	$ 0.18	0.43 %

Notes: The first three rows show the OLS coefficient, SE, and R2 of welfare loss regressed on the Consistency Index.