NBER WORKING PAPER SERIES

RELATING COGNITIVE SKILLS AND PERSONALITY TRAITS TO ECONOMIC PREFERENCES: DECISION-MAKING UNDER COGNITIVE SHOCKS AND GUESSING

Cesar Chavez Padilla
Shuaizhang Feng
James J. Heckman
Zhe Yang

Working Paper 35526
http://www.nber.org/papers/w35526

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2026

This research was supported by the Eunice Kennedy Shriver National Institute Of Child Health & Human Development of the National Institutes of Health under Award Number R01HD107079 and the U.S. National Science Foundation (NSF) under award number 2201888. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the National Science Foundation, or the National Bureau of Economic Research. We thank Joe Altonji for helpful comments on the first draft. We also thank Glenn Harrison, Patrick Kyllonen, Tamara McGavock, John Rust and Michael Woodford for comments on various drafts of the paper. Ariyan Mishra and Arjun Pathy provided helpful research assistance. Ariyan Mishra also made very helpful comments on various drafts.

Relating Cognitive Skills and Personality Traits to Economic Preferences: Decision-Making under Cognitive Shocks and Guessing
Cesar Chavez Padilla, Shuaizhang Feng, James J. Heckman, and Zhe Yang
NBER Working Paper No. 35526
July 2026
JEL No. D12, D90, I21, J24

ABSTRACT

Using rich data on Chinese schoolchildren, we study the relationships of IQ and personality traits with children's risk preferences and how they affect the precision with which those preferences are measured. Our structural model separates preferences from deliberation noise and guessing, and relates guessing to deliberation noise. Psychological measures have weak direct influences on preferences but strongly affect deliberation noise and guessing behavior. Shocks to choice have persistent individual-specific components and are not independent. Socioeconomic background has little direct influence on preferences or deliberation noise once traits are accounted for. Background variables affect psychological traits, however. Risk preferences differ across and within genders. For small gambles, girls are less risk averse than boys, but the opposite holds for larger gambles.

Cesar Chavez Padilla
The University of Chicago
Harris School of Public Policy
cesarchavezp@uchicago.edu

Shuaizhang Feng
Jinan University
Department of Economics
shuaizhang.feng@gmail.com

James J. Heckman
The University of Chicago
and IZA
and also NBER
jjh@uchicago.edu

Zhe Yang
Liaoning University
China Economic Research Institute
zhey@uchicago.edu

An online appendix is available at http://www.nber.org/data-appendix/w35526

1 Introduction

The fields of personality psychology and economics developed along independent paths, but have recently begun to converge. A substantial body of evidence now shows that traits other than cognition play important roles in predicting a wide range of behaviors. Economics has moved beyond its early focus on IQ as the primary source of achievement and flourishing lives, and now recognizes the important roles played by personality traits.[1]

This paper studies whether, to what degree, and through which channels IQ and personality traits shape the risk preferences of children. Both IQ and personality traits are associated with measures of preference as well as the consistency and quality of decision-making in laboratory and field settings. A recent literature examines the measurement of preferences and the challenge of extracting true preferences when individuals have imprecise understanding of choice problems.[2] Behavioral economists emphasize hard-wired features of the person such as limited attention, use of heuristic reasoning, and systematic mistakes as intrinsic limitations of classical rational-choice behavior (Enke et al. 2023). Standard elicitation methods, such as multiple price lists (Holt and Laury 2002), impose cognitive demands that appear to induce noise in deliberation. This literature suggests the importance of accounting for deliberation noise and elicitation mistakes in understanding choice behavior.

Early work by Borghans et al. (2008) and Almlund et al. (2011) raised the question of how personality traits relate to economic preferences. Dohmen et al. (2011) collect survey data from the German Socioeconomic Panel and correlate measures of preferences with personality traits, finding mostly weak associations. More recent contributions have advanced this literature. Andersson et al. (2016, 2020) show that the observed relationship between IQ and risk aversion partly reflects deliberation noise rather than true preference

[1] See Almlund et al. (2011) and Heckman et al. (2021) for surveys.

[2] See Sims (2003), Andersson et al. (2016), Woodford (2020), Andersson et al. (2020), and Caplin (2025). Classic studies highlight the difficulty of separating deviations from expected utility from stochastic choice error, motivating explicit noise components in structural estimation (Harless and Camerer 1994; Hey and Orme 1994).

differences. Jagelka (2024) links personality traits and preferences in a random preference model due to Thurstone (1927) that incorporates measurement error and cognitive imprecision. Accounting for measurement error and deliberation noise yields stronger and more interpretable relationships than those found in previous studies.

This paper builds on this literature and analyzes rich data on Chinese fifth graders. We elicit detailed measures of preferences and collect data on IQ, personality traits, and family background. Our paper contributes to an emerging literature on children's economic preferences initiated in work by Harbaugh et al. (2002). We extend previous work by examining the role of personality traits in deliberation noise and guessing behavior. Prior work documents substantial heterogeneity in children's risk preferences (Sutter et al. 2019; Von Gaudecker et al. 2011), but the mechanisms generating it are not systematically delineated.

We show that part of the variation in preferences reflects differences in deliberation noise rather than differences in true underlying preferences. This is consistent with previous work that shows that individuals with higher IQ tend to exhibit less noise and make more consistent choices (Andersson et al. 2016, 2020; Jagelka 2024). We find that conscientiousness and other personality traits are also associated with diminished deliberation noise, although the empirical associations differ by gender, suggesting that both cognition and personality influence the quality and stability with which preferences are expressed. We estimate two sources of decision errors: (a) deliberation noise, and (b) guessing. We show that personality traits directly affect guessing behavior and also indirectly through their impact on deliberation noise, which affects guessing. This gives a nuanced interpretation to the evidence on the role of background variables in Malmendier and Wachter (2024).

In our framework, IQ and personality traits jointly shape underlying preferences. We extend the current literature to introduce two channels for cognitive errors: deliberation noise and guessing. They mediate how these preferences translate into observed choices (see Figure 1). Deliberation noise captures imprecision in decision-making. Guessing captures responses made without deliberation. Both channels are influenced by cognitive and personality traits. Guessing also depends on the level of deliberation noise. Accounting for these diverse channels expands our understanding of the relationship between choices under risk and psychological traits.

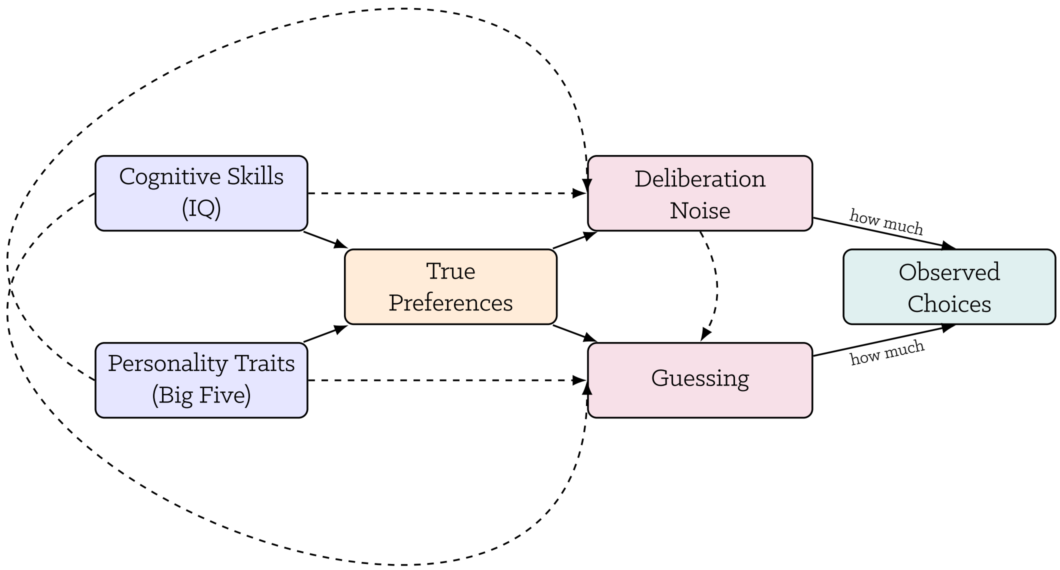


Figure 1: Schemata of Agent Responses to Choice Menus

The model developed in this paper is a structural framework that is empirically operational and novel. It can be used for analyzing many relationships between a variety of preferences and traits. Existing work on personality and preferences (Borghans et al. 2008; Dohmen et al. 2011) analyzes direct trait-preference links. We show that traits and background shape decision quality primarily through their effects on deliberation noise and guessing behavior.[3]

We extend the existing literature by examining the correlation of decision errors across lists of choices designed to elicit preferences. Independence of preference shocks across

[3] Andersson et al. (2016) show that IQ affects deliberation noise. We extend their analysis by considering a broad array of personality traits in addition to IQ and we introduce guessing as a function of these traits as well as deliberation noise.

choices is typically assumed.[4] We find high correlation in shocks that is consistent with persistent unobserved individual traits or systematic defects in decision-making across choices, contrary to the standard independence assumption in the literature. We conduct a parallel analysis for guessing behavior, allowing the variances of shocks affecting guessing to depend on personality and cognitive traits and for shocks to be correlated across items in choice lists. Here too, we find strong correlations across shocks to choices. Accounting for both types of correlations greatly improves the fit of the models and substantially changes inference about preferences.

We quantify the contributions of personality, cognition and background to: (1) true preference parameters, (2) guessing behavior, (3) shocks to preferences and guessing, and (4) persistence in shocks across choices.

We test and reject the standard CRRA specification used in most previous work on adults and children and find that the Expo-Power utility model provides a superior fit across all specifications with correlated errors, guessing, or traits.[5] The choice of functional form critically affects our estimates of risk aversion and the channels of influence depicted in Figure 1. Expo-Power functions are characterized by relative aversion parameters that depend on the level of stakes in the gambles and the endowments of children. Even fixing traits and background, there is no single risk aversion parameter, unlike the case for CRRA preferences.

We find strong gender differences in preferences and in the roles played by psychological traits in each component of both models. Environments operate primarily through the channel of psychological traits. Conditional on traits, environments have little additional direct effect on preferences, noise, or guessing.

Our findings advance the previous literature in three ways. First, we challenge the consensus that girls are more risk averse than boys, a pattern that is replicated in our data

[4] See e.g., Jagelka (2024).

[5] Our evidence on the value of Expo-Power is consistent with the work of Holt and Laury (2002), who studied young adults.

when using standard CRRA models without a correlated error structure. This disappears once we account for correlated errors and psychological traits. This stylized fact is reversed over ranges of endowments and stakes when we relax CRRA functional-form assumptions. This reversal reveals fundamental limitations in how preferences are measured using conventional methods. Properly accounting for correlated shocks even within a CRRA framework eliminates the finding of greater risk aversion for girls. Flexible functional forms combined with correlated shocks produce models over certain ranges of gambles and endowments that show that boys are more risk averse than girls, although for large endowments and gambles, girls are more risk averse.

Second, we improve on standard approaches by modeling the determinants of deliberation noise and recognizing that deliberation noise varies across individuals. When the various components of Figure 1 are modeled jointly we more accurately recover preferences. Third, we extend models of noisy cognition (Woodford 2020; Caplin 2025) by incorporating correlation across shocks, separating persistent individual differences from transitory shocks, introducing guessing as an independent source of noise, and allowing deliberation and guessing errors to be correlated. Relaxing these assumptions regarding the error structures has substantial impact on the estimation of risk preferences. An additional discovery is the importance of using teacher assessments of personality instead of widely used self-assessments by children. Model fit is better using such measures and economic estimates are more interpretable. This finding is consistent with research by Feng et al. (2022).

In this paper, we ask if preference parameters are affected by traits. This is the question posed in Borghans et al. (2008) and Falk et al. (2018). An alternative approach, following suggestions in Almlund et al. (2011), is to assess how preferences and context determine traits. We lack the data to do so effectively in our cross-sectional study. We hope to examine this question in the future using the panel data available to us. The rest of the paper is organized as follows. Section 2 presents our framework for analyzing choice under deliberation noise and guessing, extending previous models by incorporating background and personality traits and allowing shocks to preferences to influence guessing behavior. Section 3 describes our data. Section 4 introduces the CRRA and Expo-Power utility specifications.[6] Section 5 compares their fit and presents our main empirical results for the Expo-Power model. We decompose the variance in deliberation noise and guessing into persistent individual traits and transitory choice-specific shocks. We estimate how IQ and personality traits affect risk preferences, deliberation noise, and guessing behavior. Our results survive a battery of robustness checks. Section 6 quantifies the relative contributions of random transitory noise, persistent individual traits, and guessing behavior to observed mistakes in choices. Section 7 summarizes our main findings and discusses their implications for understanding measured preferences and the role of psychological traits in shaping economic decision-making during childhood.

2 Self-Knowledge and Guessing

Economic agents may not fully understand how much they truly value one alternative over another, or they may be influenced by cognitive or affective factors.[7] We model these errors as a component of a process of decision-making.

2.1 Decision under Imperfect Knowledge

Consider a binary choice between alternatives A and B. In our application to multiple price lists, the comparisons are between specific risky choices. Let U^A and U^B denote *true* utilities. Utilities may not be perfectly perceived by agents, but are subject to random

[6] See Saha (1993) and Xie (2000).

[7] A large body of research challenges the assumption that agents possess perfect self-knowledge. Behavioral economics documents that valuation and choice depend on reference points, affect, and context rather than fixed utilities (Rabin 1998; Loewenstein 2000). Recent work models cognition itself as noisy or costly, implying the presence of internal information frictions (Sims 2003; Woodford 2020; Caplin 2025). Structural and experimental evidence shows that measured decision imprecision (noise) affects revealed preferences, and that psychological traits can be related to preference parameters in models that handle decision errors (Jagelka 2024). Barseghyan et al. (2018) present a valuable summary of approaches to estimate risk preferences from non-experimental data.

cognitive disturbances τ^A and τ^B when evaluating the options. Observed choice D_A is:

$$D_A = \mathbb{1}(U^A - U^B + \tau^A - \tau^B \geq 0), \tag{1}$$

so decisions depend on both genuine preference differences and perceptual noise. For convenience, define $\varepsilon \equiv \tau^B - \tau^A$, $\mathrm{Var}(\tau^A) = \sigma^2_{\tau^A}$ and $\mathrm{Var}(\tau^B) = \sigma^2_{\tau^B}$, $\mathrm{Var}(\varepsilon) = \sigma^2_\varepsilon = \sigma^2_{\tau^A} + \sigma^2_{\tau^B} - 2\,\mathrm{Cov}(\tau^A, \tau^B)$. A positive ε favors B, and a negative one favors A. A small value of σ^2_ε implies more stable choices, while a large σ^2_ε implies near-random choices. Normalizing by σ_ε:

$$D_A = \mathbb{1}\left(\frac{U^A - U^B}{\sigma_\varepsilon} > \frac{\varepsilon}{\sigma_\varepsilon}\right). \tag{2}$$

With perfect self-knowledge ($\varepsilon = 0$), the choice is deterministic:

$$D^*_A = \mathbb{1}(U^A - U^B \geq 0). \tag{3}$$

The distinction between (2) and (3) captures decision imprecision.

2.2 Information and the Cost of Precision

Agents can reduce uncertainty by exerting effort and acquiring information. Let $\mathcal{I}_A$ and $\mathcal{I}_B$ denote information about choices available to agents through effort collected in vector $\mathcal{I} = (\mathcal{I}_A, \mathcal{I}_B)$. Thus, more effort increases precision. As $\mathcal{I} \uparrow$, $\sigma^2_j(\mathcal{I}) \downarrow$, $j \in \{A, B\}$.

This captures models of rational inattention and costly cognition in which agents choose information precision (Sims 2003; Woodford 2020; Caplin 2025). Let $C(\mathcal{I}_A)$ and $C(\mathcal{I}_B)$ be increasing, convex cost functions. Since $D_A + D_B = 1$, realized utility is $U^A D_A + U^B D_B$.[8] Agents are assumed to choose $\mathcal{I}$ to maximize expected utility net of costs:

$$E\big(\left[U^A + \tau^A\right] \mathbb{1}(U^A - U^B > \varepsilon) + \left[U^B + \tau^B\right] \mathbb{1}(U^B - U^A > -\varepsilon)\big) - C(\mathcal{I}_A) - C(\mathcal{I}_B). \tag{4}$$

[8] We ignore the possibility that agents make no decision, and instead assume that they guess.

Agents may be at interior solutions, equating marginal benefits to marginal costs, or may go to corners and simply guess if the marginal benefit of effort on a particular choice is too small.

If $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, define $q \equiv \varepsilon/\sigma_\varepsilon \sim \mathcal{N}(0,1)$. The expected payoff (ignoring costs) can then be expressed using $\Phi(\cdot)$, the standard normal CDF. When noise gets higher, $\sigma_\varepsilon(\mathcal{I}) \to \infty$ and choices approach 50–50 guessing. Thus, $\Pr(D_A = 1) = \Phi\left(\frac{U_A - U_B}{\sigma_\varepsilon}\right)$, which tends to $\frac{1}{2}$ as $\sigma_\varepsilon \to \infty$, provided U_A and U_B are bounded. If choices are similar, so that $U_A \to U_B$ for any value of σ_ε, the choice probability is also $\frac{1}{2}$. We obtain the same result for any random variable centered at zero. By the principle of insufficient reason, if the two options in a binary choice become more similar, agents assign a probability of $\frac{1}{2}$ to each, just like the case of guessing.

2.3 Stochastic Specifications

Cognitive noise varies across choices and individuals. We model the noise for individual i in choice j among J choices as

$$\varepsilon_{ij}^D = \lambda_j \, \theta_i^D + u_{ij}^D, \qquad i = 1, \dots, I, \tag{5}$$

where θ_i^D is a persistent (trait-like) component and u_{ij}^D is a transitory shock, and I is the sample size. We assume

$$\theta_i^D \perp\!\!\!\perp u_{ij}^D, \qquad u_{ij}^D \sim \mathcal{N}\left(0, \sigma^2_{u_{ij}^D}\right), \tag{6}$$

where the idiosyncratic variance $\sigma^2_{u_{ij}^D}$ may depend on observable factors such as background and psychological traits. This introduces heteroskedasticity in the precision of decision-making across individuals and choices, a central feature of one version of the pioneering Thurstone (1927) model. Assuming exchangeability among choices faced by agents, we set $\lambda_j \equiv \lambda$, and normalize $\lambda = 1$. We parameterize $\sigma^2_{u_{ij}^D}$ and $\sigma^2_{\theta^D}$ to depend on psychological traits and background variables for each $i = 1, \dots, I$.

2.3.1 Guessing in Decision Making

If effort in decision-making is not exerted, or if choices are too similar, the individual guesses.[9] If it is exerted and choices are distinct, decisions may follow a utility-based rule, albeit with cognitive imprecision, as described in Woodford (2020), Belzil and Jagelka (2024), and Caplin (2025). The error structure for the guessing shock ε_{ij}^G for person i on choice j parallels that for the decision-noise shock in equation (5):

$$\varepsilon_{ij}^G = \omega_j \, \theta_i^G + u_{ij}^G, \qquad \theta_i^G \perp\!\!\!\perp u_{ij}^G, \quad u_{ij}^G \sim \mathcal{N}(0, \sigma^2_{u_{ij}^G}), \tag{7}$$

where θ_i^G represents a persistent (trait-like) guessing tendency, and the idiosyncratic component u_{ij}^G has variance $\sigma^2_{u_{ij}^G}$ that may depend on observable characteristics. We assume exchangeability across tasks and normalize $\omega_j = \omega = 1$. The precision of engagement or guessing behavior is permitted to vary across individuals and tasks.

Let $G_{ij} = 1$ indicate guessing on task j:

$$G_{ij} = \mathbb{1}\Big(\eta(\sigma^2_{u_i^D}, \boldsymbol{X}_i) + \omega_j \, \theta_i^G + u_{ij}^G > 0\Big), \tag{8}$$

where $\eta(\sigma^2_{u_i^D}, \boldsymbol{X}_i)$ is a function of observed variables capturing how guessing depends on psychological traits, deliberation noise, and background variables. We consider three parameterizations of the guessing probability: (i) one in which guessing depends directly on IQ and personality traits, (ii) a second in which deliberation noise σ_{ij}^D affects guessing decisions, with IQ and personality traits influencing guessing only indirectly through deliberation noise, and (iii) a third specification that incorporates both channels as described in the previous two models. We use the first specification as the default in our full model

[9] This guessing component plays the same role as the guessing parameter in the three-parameter logistic (3PL) item response model, where a respondent may answer correctly by chance rather than through ability, see Lord and Novick (1968).

and compare it with the alternatives. We assume that the shocks u_{ij}^D and u_{ij}^G are independent for all (i, j) and also across persons, and that both ε_{ij}^D and ε_{ij}^G exhibit within-choice correlation structures as previously described.

2.3.2 Causality

We assume $\boldsymbol{X}$ is exogenous. This justifies causal interpretation of the estimates reported in this paper.

2.4 Probabilities of Choice

Agents make binary choices across items. Let $Y_{ij} \in \{A, B\}$ denote the observed choice for agent i on choice j. The probability of B conditional on X_i (implicit to simplify notation) is

$$\Pr(Y_{ij} = B) = \underbrace{\Pr(Y_{ij} = B \mid G_{ij} = 0)}_{\text{deliberate}} \underbrace{\Pr(G_{ij} = 0)}_{\text{don't guess}} + \tfrac{1}{2} \underbrace{\Pr(G_{ij} = 1)}_{\text{guess}}. \tag{9}$$

We assume that

$$\mathrm{Var}[(\varepsilon_{i1}^D \dots \varepsilon_{iJ}^D)'] = \sigma_{\theta^D}^2 \, (\boldsymbol{i}_J \boldsymbol{i}'_J) + \sigma_{u^D}^2 \, \boldsymbol{I_{JJ}}, \tag{10}$$

where $\boldsymbol{i}$ is a $J \times 1$ column vector of ones, with intra-class correlation parameter

$$ICC^D = \frac{\sigma_{\theta^D}^2}{\sigma_{\theta^D}^2 + \sigma_{u^D}^2}. \tag{11}$$

We make a parallel assumption for guessing shocks,

$$\mathrm{Var}[(\varepsilon_{i1}^G \dots \varepsilon_{iJ}^G)'] = \sigma_{\theta^G}^2 \, (\boldsymbol{i}_J \boldsymbol{i}'_J) + \sigma_{u^G}^2 \, \boldsymbol{I_{JJ}}, \tag{12}$$

with

$$ICC^G = \frac{\sigma_{\theta^G}^2}{\sigma_{\theta^G}^2 + \sigma_{u^G}^2}. \tag{13}$$

The log-likelihood of the model and an identification analysis are presented in Appendix A. We assume independence across agents. The variables θ^D and θ^G are integrated out in forming the choice probabilities. We allow θ^D and θ^G to be correlated and parameterize this correlation, alone with the idiosyncratic and persistent variances $\sigma^2_{u^D}$, $\sigma^2_{u^G}$, $\sigma^2_{\theta^D}$, $\sigma^2_{\theta^G}$, as functions of individual background, IQ, and personality traits, while assuming that the elements of u^D and u^G are mutually uncorrelated. These heteroskedastic structures capture how unobserved behavioral noise varies across individuals and tasks. It is possible to model group-level or school-level effects as additional random clusters, but we do not do so in this paper.

3 The Mianzhu Data

We use the *Longitudinal Study of Children's Development in Mianzhu*, a large-scale study of Chinese schoolchildren following students from Grade 4 through high school graduation. The cohort spans rural and urban schools in Mianzhu, Sichuan Province, capturing approximately 7,200 students across 18 schools with wide variation in socioeconomic conditions, family structure, and educational context.

The study has three features important to our analysis. First, we simultaneously measure economic preferences (risk, ambiguity, time, altruism, reciprocity), IQ, and personality traits (Big Five) in a large and socioeconomically diverse population. Second, we perform elicitations from multiple price lists allowing us to separate persistent traits from transitory shocks. Third, we collect personality assessments from students, teachers, and parents, using teacher reports and thus mitigating self-report bias that often contaminates personality–behavior relationships.[10]

In this paper, we use the risk preferences of 2,204 fifth-grade students (1,115 boys and 1,089 girls) surveyed in 2019 who completed the IQ assessment, participated in all risk preference elicitations, and whose personality traits were evaluated by homeroom teachers

[10] For context on child/adolescent preference measurement and samples, see Sutter et al. (2019). Examples with repeated, incentivized elicitations include Alan and Ertac (2018). On multi-informant psychological assessments and measuring personality traits, see Kautz and Zanoni (2014) and Feng et al. (2022).

who have observed students over multiple grades. Although we have panel data on children through college, we focus on the 2019 Grade-5 wave, the wave in which risk preferences, IQ, and teacher-reported personality were measured together. Risk preferences were elicited using six incentivized multiple price list (MPL) tasks administered via tablet computers, presented through child-friendly game interfaces with real payoffs. Most tasks offer a sequence of choices between a safe option and a lottery, and a subject's risk attitude is identified from the point at which they switch between the two. Some tasks involve choices between two lotteries. The full protocol and the six choice tables are given in Appendix B. Students' IQ was measured using the 60-item Raven's Standard Progressive Matrices, a widely used non-verbal test of fluid intelligence, with raw scores standardized using the pooled (boys and girls) sample mean and standard deviation within each grade level. Personality traits were assessed by teachers using a subset of the Big Five Inventory-2 (BFI-2), consisting of 20 items (4 per dimension) that capture conscientiousness, agreeableness, extraversion, openness to experience, and emotional stability (the inverse of neuroticism), based on a validated Chinese translation. Personality traits were also assessed by students using a longer version of the BFI-2 consisting of 62 items. Complete details on IQ measurement, personality traits, and their psychometric properties are provided in Appendix C.

Variation in parental migration and household composition provides heterogeneity in family inputs. The survey tracks patterns of each child's left-behind status semester-by-semester from Grade 1 through the first semester of Grade 5 (nine semesters), measuring whether each parent migrated for work for at least three months while returning home no more than once weekly. We construct two measures: (1) for each left-behind category (father only, mother only, or both parents), the proportion of semesters the child experienced that family status, and (2) the Left-Behind Volatility Index, which counts transitions in migration status, capturing household instability relevant to personality and preference formation (Hill et al. 2001; Fomby and Cherlin 2007).

3.1 Sample Characteristics and Measurement

Feng et al. (2022) show the greater predictive power of teacher-reported attributes (at least for academic achievement). We use teacher-reported Big Five personality traits rather than children's self-reports. Teachers observe children across multiple contexts over extended periods, potentially providing more objective assessments than self-reports, which may be influenced by transitory mood states or social desirability bias.[11] While the noise and guessing parameters (σ_{u^D}, ICC^D, ICC^G, and the guessing probability) are stable within about 5–10% across measurement sources, the deliberation–guessing correlation ρ is more sensitive (teacher -0.15 vs. self -0.19 for boys, and 0.11 vs. 0.16 for girls). Teacher-report coefficients are generally more precisely estimated, with similar risk-aversion estimates across sources. Appendix figure C.1 illustrates systematic divergence between teacher and student self-reports across grades, with self-reports higher for Openness and Extraversion, while for Conscientiousness and Agreeableness teacher reports start higher and self-reports rise to cross them by around Grade 5.

Table 1 shows that substantial gender differences appear in personality traits. Girls demonstrate markedly higher Conscientiousness (0.631 vs. 0.246) and Agreeableness (0.837 vs. 0.581)—patterns that are consistent with the developmental psychology literature and persist across Grades 4–6 (Appendix figure C.5). Extraversion and Openness levels are comparable between genders, while girls hold a sizable advantage in Emotional Stability. Raw scores on IQ are essentially identical across boys and girls. Demographics, parental education, and family structure and migration status are reported by gender in Appendix table E.1.

3.2 Correlations Among Cognitive and Personality Measures

Appendix tables E.2, E.3, and E.4 show moderate positive correlations among cognitive and non-cognitive measures. IQ correlates positively with all Big Five traits, strongest with

[11] Appendix D validates this choice.

Table 1: Descriptive Statistics: Grade 5 in 2019

	Boys		**Girls**		**Difference**	
	Mean	S.E.	Mean	S.E.	Mean	S.E.
Personality traits						
Conscientiousness	0.246	0.021	0.631	0.020	0.385***	0.029
Agreeableness	0.581	0.026	0.837	0.024	0.256***	0.036
Extraversion	0.066	0.023	0.064	0.025	−0.002	0.034
Openness to Experiences	0.437	0.020	0.454	0.020	0.017	0.028
Emotional Stability[a]	0.282	0.022	0.461	0.021	0.179***	0.031
Observations	1,115		1,089		2,204	

Notes: Descriptive statistics for Grade 5 students surveyed in 2019. Columns report means with standard errors (S.E.) in the adjacent column. The Difference column reports the girls-minus-boys mean difference, with significance from a two-sample test of equality of means. Big Five traits are predicted using the factor model in Heckman et al. (2013), with estimation-error corrections via the Bartlett GLS estimator. For each trait, the sample mean is normalized by fixing the "Neutral" response to the first item to zero, and the standard deviation is normalized by fixing the factor loading of the first item to one. Full descriptive statistics, including demographics, parental education, and family structure and migration status, appear in Appendix table E.1. [a]This trait, when reverse-coded, is also known as Neuroticism in the Big Five Inventory. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Openness ($r = 0.25$). Among personality traits, Conscientiousness shows strong correlations with Agreeableness ($r = 0.64$) and Emotional Stability ($r = 0.56$). The correlation structure is largely consistent across gender, though girls exhibit somewhat stronger correlations among the Big Five. These modest correlations suggest that IQ and personality traits capture distinct sources of individual heterogeneity relevant for decision-making.

3.3 Socioeconomic Gradients and Environmental Influences

The sample displays considerable socioeconomic diversity. Approximately 23% of students live in urban areas, fewer than one-third of fathers and one-quarter of mothers completed high school, and about 37% of boys and 41% of girls have at least one sibling. A substantial share of children experience parental migration for work during the nine semesters since Grade 1. On average per semester, about 25%, 7%, and 17% of boys are left behind by their father only, mother only, and both parents, respectively. Girls exhibit similar patterns across these categories.

Appendix figures C.2 and C.3 document strong socioeconomic gradients in cognitive development. Students whose parents have a college education score approximately 0.5 standard deviations higher on IQ tests by Grade 6 than those whose parents have only primary education. Appendix figure C.4 demonstrates strong urban-rural differences in IQ.

Appendix tables E.5 and E.6 examine how family background and migration status predict the individual characteristics reported in Table 1. Urban residence has mixed effects: urban boys score higher on IQ and Agreeableness but lower on Conscientiousness, while urban girls show substantially lower scores across most personality dimensions—patterns that differ from unconditional comparisons in Appendix figure C.4.

Parental education shows positive effects. Father's high school education is significantly associated with greater Conscientiousness, Agreeableness, and Openness for boys, with similar patterns for girls, consistent with developmental gradients in Appendix figures C.2 and C.3.

Parental migration effects display gender-specific patterns. For boys, current left-behind status shows limited direct effects on traits, but left-behind volatility significantly reduces Openness, Extraversion and IQ. For girls, being left behind by the mother is associated with lower Conscientiousness, Extraversion and Openness, while left-behind volatility is associated with lower IQ.

These relationships between family background and individual traits provide important context for interpreting structural estimates. In Section 5, we demonstrate that differences in risk preferences and decision-making quality across demographic groups operate mainly through their effects on IQ and personality traits, which in turn shape both true preferences and the precision of decision-making[12], as well as the probability of guessing. Environments

[12] Appendix F demonstrates that cognitive/personality traits and socioeconomic background capture largely independent sources of heterogeneity: the level parameters—r, σ_{u^D}, and ICC^D—are stable across the SES-only, traits-only, and combined specifications (within about 10%), while the guessing probability and the deliberation–guessing correlation ρ are more sensitive to whether traits are included (e.g., boys' ρ moves from -0.22 with SES alone to -0.15 with traits, and girls' guessing probability from 0.16 to 0.19), consistent with our main finding that traits operate primarily through the guessing and correlation

mainly affect personality traits, not preferences.

3.4 Some Evidence on Decision Consistency

Before presenting our structural models and their estimates, we examine simple statistical evidence on decision consistency. Appendix Table E.8 presents linear probability models in which the dependent variable measures switching behavior in multiple price lists, defined as oscillating between the two options across successive items as the incentives increasingly favor one option. This switching behavior is a source of identifiability of deliberation noise and guessing for our model.

Several patterns foreshadow our structural findings. IQ significantly reduces inconsistent behavior for both boys and girls. Conscientiousness reduces inconsistencies for girls, while Openness strongly reduces them for boys. Interestingly, Extraversion increases inconsistencies for boys but reduces them for girls, hinting at gender-specific attention mechanisms. Father's high school education significantly reduces inconsistencies for girls, while the Left-Behind Volatility Index marginally increases inconsistent behavior for boys, suggesting household instability may reduce decision quality. These patterns motivate our structural approach, which disentangles how personality traits and background operate through true risk preferences, cognitive precision in deliberation, guessing behavior, and serial correlation in preference shocks.

4 Specification of Utility Functions: CRRA versus Expo-Power

A traditional point of departure is the constant relative risk aversion (CRRA) utility function, used in Pratt (1964), Arrow (1965), Harrison et al. (2007), Andersen et al. (2008), Andersson et al. (2016), Sutter et al. (2019), Andersson et al. (2020), and Belzil and Jagelka

channels.

(2024):

$$U_{\text{CRRA}}(s+w) = \begin{cases} \dfrac{(s+w)^{1-r}-1}{1-r}, & r \neq 1, \\ \\ \ln(s+w), & r = 1, \end{cases} \tag{14}$$

where s denotes the payoff from an experimental choice, w denotes reference wealth, and r is a risk aversion parameter. CRRA exhibits constant relative risk aversion:

$$\text{RRA}(s+w) = r, \tag{15}$$

independently of wealth. We parameterize r to depend on psychological traits.

A more flexible specification that nests CRRA was introduced by Saha (1993) and Xie (2000), and used by Holt and Laury (2002). The *Expo-Power* utility function is:

$$U_{\text{ExpoPower}}(s+w) = \frac{1}{\beta}\left[1 - \exp\left(-\beta\,\frac{(s+w)^{1-r}-1}{1-r}\right)\right]. \tag{16}$$

This specification nests both CRRA ($\beta \to 0$) and CARA ($r = 0$), and produces a rich characterization of risk parameters that are wealth-dependent:

$$\text{RRA}(s+w) = r + \beta\,(s+w)^{1-r}. \tag{17}$$

The parameter r governs risk aversion curvature, while β scales the sensitivity of risk aversion to the level of $(s+w)$.[13]

The Expo-Power specification allows risk aversion to vary with total resources. When $r < 1$, individuals exhibit increasing relative risk aversion (IRRA) as $(s+w)$ increases if $\beta > 0$, and decreasing relative risk aversion (DRRA) if $\beta < 0$. Our estimates support IRRA and rule out DRRA. The Expo-Power specification also bounds utility from above when $\beta > 0$, introducing a natural notion of satiation. Additionally, it exhibits desirable

[13] For both the CRRA and Expo-Power models, we assume that reference wealth is fixed at $w = 1$. In Appendix G, we relax this assumption by fixing w at alternative values. In Appendix H, we relax it further by estimating w as a free parameter and allowing it to vary with children's socioeconomic status.

concavity properties under the condition $\beta(s+w)^{1-r} > -r$.

This flexibility allows observed heterogeneity in choices to reflect differences in wealth, and stakes rather than fixed curvature alone (Xie 2000).[14] For 10- and 11-year-olds, wealth-dependent risk aversion captures how stake magnitude affects decision-making. Payoffs range from 1 to 200 tokens across the six lists, a range that determines prize tier achievement. At low stakes (1 token), individual choices minimally affect cumulative earnings and final prizes. At high stakes (200 tokens), a single decision substantially influences outcomes. The Expo-Power specification allows $\mathrm{RRA}(s+w)$ to vary with this wealth level, capturing psychological shifts across the stake range rather than imposing CRRA's rigid constant-risk-aversion assumption. This flexibility is essential for children whose preference structures remain in development and who may process small versus large gambles qualitatively differently. The interpretation of w as a reference wealth level is well defined for adults. For children, it is less so. We conduct a robustness analysis varying w. Our main results are not much affected (Appendix G). Chávez et al. (2026) consider a wider class of models of risk aversion for children. They formally model w and explicitly incorporate family background. That paper examines a richer class of child preferences including those considered by Harbaugh et al. (2002).

5 Empirical Results

Before turning to our parameter estimates, it is helpful to summarize our general findings. We then document these claims. See Table 2 for an overview of the fit of alternative models. The table reports log-likelihoods and likelihood-ratio tests (Panel A) and the AIC (Akaike Information Criterion) (Panel B) across specifications for the CRRA and Expo-Power models. Each component improves fit over the baseline, with the ICC structure

[14] A related but more restrictive form appears in Saha (1993), who studies an exponential–CRRA hybrid for agricultural decision making (see Appendix I for a detailed discussion). Holt and Laury (2002) show that this specification is more concordant with data on risky choices than CRRA.

contributing the most. Expo-Power is favored by the AIC[15] in the full model, and implied risk aversion varies with which components are included. The table is discussed in detail in Appendix J. We introduce psychological measures as arguments in the structural parameters of the CRRA and Expo-Power models, and consider how psychological variables modify the estimates of the various parameters of each model.

Table 2: Comparing CRRA vs. Expo-Power Across Different Specifications

	Boys			Girls		
Specification	CRRA	Expo-Power	LR Test	CRRA	Expo-Power	LR Test
Panel A: Log-Likelihood						
Baseline (None)	−36015.707	−36014.707	2.0	−36012.605	−35935.292	154.626***
			(0.157)			(0.000)
Only IQ and Traits	−35663.938	−35632.265	63.346***	−35743.955	−35646.788	194.334***
			(0.000)			(0.000)
Only ICC	−34059.244	−34051.104	16.28***	−34244.707	−34153.581	182.252***
			(0.000)			(0.000)
Only Guessing	−35990.429	−35985.982	8.894***	−36010.794	−35921.558	178.472***
			(0.003)			(0.000)
Full Model	−32777.287	−32754.608	45.358***	−33333.503	−33232.961	201.084***
			(0.000)			(0.000)
Panel B: Akaike Information Criterion (AIC)						
Baseline (None)	72035.414	72035.414		72029.21	71876.584	
Only IQ and Traits	71355.876	71306.53		71515.91	71335.576	
Only ICC	68124.488	68110.208		68495.414	68315.162	
Only Guessing	71986.858	71979.964		72027.588	71851.116	
Full Model	65638.574	65607.216		66751.006	66563.922	

Notes: p-values in parentheses for Panel A. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Panel A presents log-likelihood values and likelihood ratio test statistics $\lambda_{LR} = 2(\mathcal{L}_{\text{Expo}} - \mathcal{L}_{\text{CRRA}})$. Panel B shows the Akaike Information Criterion, where lower values indicate better fit. "Baseline (None)" is the simplest model without IQ/traits, ICC, or guessing.

Our major findings are as follows:

(1) Preferences differ by gender.

[15]We separately test the null hypotheses of an independent error structure and the absence of guessing behavior using likelihood ratio tests. Due to boundary constraints in the parameter space for ICC and guessing, the log-likelihood ratios associated with these null hypotheses follow a mixed χ^2 distribution: a point mass at 0 with probability 1/2 and a χ^2 distribution with 1 degree of freedom with probability 1/2. See Chernoff (1954) and Self and Liang (1987). We approximate this mixed χ^2 distribution to calculate the p-values for these two likelihood ratio tests.

(2) For boys, CRRA and Expo-Power fit equally well in terms of AIC in specifications with independent errors across choices, without guessing, and without psychological or background variables. When psychological and background variables are added alone or correlated decision shocks alone, Expo-Power fits better. For girls, Expo-Power fits better in all cases.

(3) For both genders and both models, adjusted fit[16] is better when the psychological variables are added. Their addition does not substantially change the estimated parameters computed from average predictions over the sample. There are sharp gender differences in the predictive value of psychological traits.

(4) Shocks are strongly correlated across the items of a multiple price list elicitation. This finding challenges the current practice of assuming that shocks are independent across choices. Accounting for correlated shocks to true preferences and for guessing greatly affects the performance of the model. Correlated shocks are an important feature of all specifications of the models for both genders.

(5) Introducing guessing improves fit. Disaggregating cognitive errors into guessing and noise is a contribution of this paper. When guessing is introduced, the role of deliberation noise in decision errors is reduced in all specifications for both genders. Eliminating guessing has only a small effect on the estimated structural preference parameters, operating mainly through β. It strengthens the role of deliberation noise and increases correlation in shocks. Guessing makes a substantial contribution to utility loss due to choice mistakes and is relatively more important than deliberation noise and correlated errors. Point estimates suggest that cognitive noise variance increases guessing for girls but not for boys, though the effect weakens when traits are added and is not statistically distinguishable from zero.

(6) Background variables do not directly affect economic preference parameters, but do affect the psychological measures that affect noise and guessing.

(7) Under baseline CRRA, without modeling correlated shocks across choices, boys

[16] Adjusted for introduction of additional variables, as in AIC.

appear less risk averse than girls (girls' single r is 12% higher, 0.95 vs. 0.85). In the full CRRA model that adds correlated shocks, guessing, and traits, the estimates converge and the gender disparity disappears (both $r \approx 0.97$). Under Expo-Power, relative risk aversion depends on s and w. At low stakes and low wealth, girls are less risk averse than boys, while at higher stakes and wealth the ordering reverses (see Figure 2). Modeling the ICC structure raises the estimated level of relative risk aversion for both genders, with the largest upward shift for boys. Ignoring correlated deliberation noise therefore understates how risk averse children are, particularly boys.

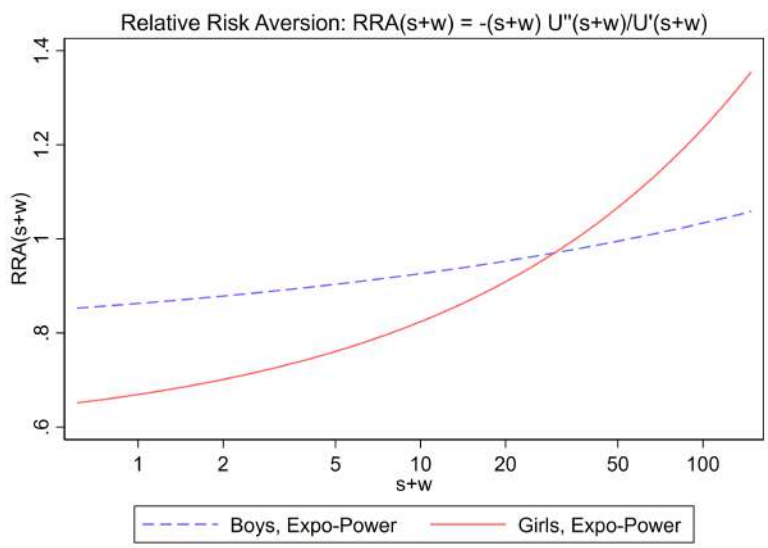


Figure 2: Relative Risk Aversion as a Function of Total Wealth by Gender (Full Expo-Power Model)

Notes: The figure plots the coefficient of relative risk aversion as a function of total wealth, $RRA(s+w) = -(s+w)\frac{U''(s+w)}{U'(s+w)} = r + \beta(s+w)^{1-r}$, separately for boys (dashed blue curve) and girls (solid red curve). The parameters r and β are evaluated at the sample means of the IQ and personality variables.

We present a full analysis of each specification. Our analysis for CRRA is given in the Appendix. That for Expo-Power is presented in the text. This section proceeds as follows. Section 5.1 sets out how traits enter the models. Section 5.2 presents the contributions of model components to the full model[17], adding correlated shocks, ICC structure, and guessing, and compares the CRRA and Expo-Power specifications. Section 5.3 asks whether IQ and personality traits matter for risk preferences, deliberation noise, and guessing,[18] and Section 5.4 compares alternative guessing specifications. Section 5.5 consolidates our main findings on the gender gap. Section 5.6 examines heterogeneity in risk preferences, and Section 5.7 reports a battery of robustness checks.[19]

5.1 Parameterization of the Model

Before presenting full model results, we describe how IQ and personality traits enter the structural parameters. This discussion facilitates understanding of the coefficient estimates reported below.

5.1.1 Parameters and Their Dependence on Observables

The Expo-Power model involves seven key parameters, each of which we allow to depend on individual characteristics $\boldsymbol{X}_i$ (IQ, personality traits, and background variables). We adopt similar parameterizations for the CRRA model.

Risk aversion curvature $r(\boldsymbol{X}_i)$ captures the curvature of the utility function, where higher r implies greater risk aversion. The wealth-dependence $\beta(\boldsymbol{X}_i)$ is the Expo-Power parameter governing how risk aversion varies with wealth w. When $\beta > 0$ and $r < 1$, or when $\beta < 0$ and $r > 1$, individuals exhibit increasing relative risk aversion as wealth increases (for CRRA, $\beta = 0$ by construction). Transitory noise $\sigma_{u^D}(\boldsymbol{X}_i)$ is the standard deviation of task-specific deliberation noise, where lower values indicate more precise decision-making. $\text{ICC}^D(\boldsymbol{X}_i)$ is the intra-class correlation for deliberation noise, capturing the proportion

[17] We estimate a baseline CRRA model, reported in Appendix K, adding correlated shocks, ICC structure, and guessing. It compares the CRRA and Expo-Power specifications. The role of the ICC structure is examined in Appendices L and M, the role of guessing in Appendix N, and the full-model estimates appear in Appendix O.

[18] The CRRA results are compared with Expo-Power in Appendix P.

[19] The CRRA results are compared with Expo-Power in Appendices F, G, and H.

of total noise variance attributable to persistent individual differences. $\text{ICC}^G(\boldsymbol{X}_i)$ is the intra-class correlation for guessing behavior, capturing persistent individual differences in the propensity to guess. The correlation $\rho(\boldsymbol{X}_i) = \text{Corr}(\theta^D, \theta^G)$ captures the relationship between persistent components of deliberation noise and guessing propensity—this parameter is particularly important as it reveals whether individuals with traits favoring choices with greater payoff variance tend to guess more or less often. Finally, guessing probability $\Pr(G_{ij} = 1 \mid \boldsymbol{X}_i)$ is the probability of guessing rather than deliberating on a given task.

In model estimation, we re-standardize the five personality traits in $\boldsymbol{X}_i$ using the pooled-gender sample of Grade 5 students in 2019. Accordingly, all marginal effects of the personality traits are interpreted as the effect of a one-standard-deviation increase in each of the five domains.

5.1.2 Parameter Dependence on X

Each parameter is specified as a function of observables. For parameters that are naturally unbounded (such as r and β), we use a linear model:

$$\pi(\boldsymbol{X}_i) = \sum_{k=0}^{K} \pi_k \cdot X_{ki}, \tag{18}$$

where π_k represents parameters of interest and $\pi(\boldsymbol{X}_i)$ represents a generic parameter.

For parameters that are positive (such as the variance $\sigma^2_{u^D}$) or bounded between 0 and 1 (such as ICC and guessing probability), we apply appropriate transformations. For variances, we specify:

$$\sigma^2_{u^D}(\boldsymbol{X}_i) = \exp\left(\sum_{k=1}^{K} \gamma_k \cdot X_{ki}\right), \tag{19}$$

where γ_k represents the parameters of interest. For ICC parameters and guessing probabilities, we use logistic transformations to ensure values remain in $(0, 1)$. For the correlation parameter ρ, we use a hyperbolic tangent transformation to ensure values remain in $(-1, 1)$.

5.2 Components of the Full Model

In this section, we present the analysis of a stripped-down baseline model as well as that of the full model. We discuss the components of the best-fitting model. In Appendix J, we incrementally add model components to both the CRRA and Expo-Power models to understand their contributions to fit singly and jointly.

5.2.1 Baseline Model Comparison: CRRA vs. Expo-Power

Table 3 compares CRRA and Expo-Power in their simplest form—without IQ or personality traits, ICC structure, or guessing behavior. The results reveal striking gender differences. For girls, Expo-Power dramatically outperforms CRRA (LR = 154.6, $p < 0.001$), with risk aversion curvature r some 42% lower than under CRRA (0.55 vs. 0.95). For boys, the baseline comparison shows no significant difference (LR = 2.0, $p = 0.157$). In AIC terms the two baseline models for boys are effectively tied, since the one-point log-likelihood gain from Expo-Power offsets its additional parameter, so the criterion does not discriminate at baseline. Expo-Power's advantage for boys emerges once the ICC structure, guessing, and covariates are modeled (Table 2).

Table 3: Baseline Model Comparison: CRRA vs. Expo-Power

	Boys		Girls	
	CRRA	Expo-Power	CRRA	Expo-Power
	(1)	(2)	(3)	(4)
r in Expo-Power + CRRA	0.845^{***}	0.767^{***}	0.946^{***}	0.546^{***}
	(0.014)	(0.055)	(0.013)	(0.025)
β in Expo-Power		0.034^{*}		0.080^{***}
		(0.018)		(0.004)
Deliberation Noise s.d. (σ_{u^D})	1.368^{***}	1.485^{***}	1.056^{***}	2.013^{***}
	(0.073)	(0.128)	(0.053)	(0.163)
Loglikelihood	−36015.707	−36014.707	−36012.605	−35935.292
λ_{LR}: Expo-Power vs. CRRA	2.000		154.626	
p-value of LR Test	[0.157]		[0.000]	

Notes: We compute predictions for each observation and average them over the sample. Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Maximum likelihood estimates are based on models that exclude IQ and personality traits, ICC structure, and guessing behavior.

5.2.2 The Full Model

We next present the full model specification that incorporates: (1) IQ and Big Five traits that affect all parameters, (2) a correlated error structure that captures persistent heterogeneity in the unobservables affecting deliberation, and (3) explicit guessing behavior with correlation of deliberation shocks and guessing shocks. This is our preferred model. When background variables are added, they barely change the estimates of the economic model. Any effect of background variables on choice parameters operates through psychological variables. See Appendix F.

Appendix table J.2 traces how the preference parameters change as we move from the baseline specification (CRRA without ICC, guessing, or traits) to the full model. We report preferences in terms of the underlying risk aversion curvature r and, for Expo-Power, the wealth-dependence parameter β. For CRRA, r is the single risk-aversion parameter and $\beta = 0$ by construction. For Expo-Power, r is the risk aversion curvature and β governs how relative risk aversion changes with wealth. The comparison reveals two distinct channels through which modeling choices affect gender gaps in measured risk aversion. Complete coefficient estimates appear in Appendix tables O.1–O.2 in Appendix O.

The sensitivity of estimates to introducing correlation in preference shocks across choices is a major finding of this paper. Figure 3 modifies Figure 2 to compare relative risk aversion for the best-fitting Expo-Power model with and without ICC^D. In the absence of correlated shocks, the wealth/stakes switch-over point occurs at a substantially lower level of total wealth than once the ICC structure is included. This implies that, without correlated shocks, girls appear more risk averse than boys over a wider range of stakes and wealth.

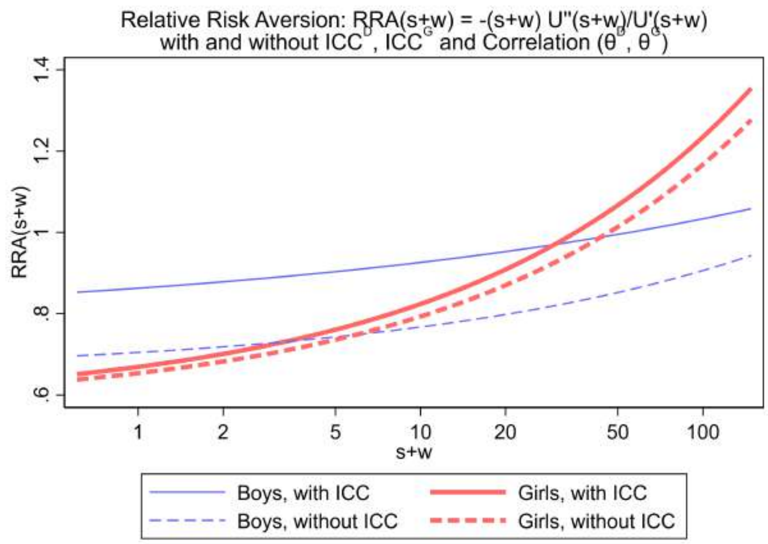


Figure 3: Relative Risk Aversion as a Function of Total Wealth, With and Without the ICC Structure (Full Expo-Power Model)

Notes: The figure plots the coefficient of relative risk aversion as a function of total wealth, $RRA(s+w) = -(s+w)\frac{U''(s+w)}{U'(s+w)} = r + \beta(s+w)^{1-r}$, separately for boys (thin blue curve) and girls (thick red curve), estimated with the ICC structure (solid lines) and without it (dashed lines). The parameters r and β are evaluated at the sample means of the IQ and personality variables. Modeling the ICC structure increases the estimated level of relative risk aversion for both genders, with the largest upward shift observed for boys. It also shifts the wealth level at which the boys' and girls' profiles cross to a substantially higher level of wealth.

5.2.3 Impacts on Estimated Model Parameters

Table 4 examines the best-fitting Expo-Power functional form by itself and demonstrates how the sequential inclusion of each error-structure component—ICC and guessing—affects the parameter estimates. For boys, risk aversion curvature r is essentially flat across specifications (from 0.767 to 0.760, 0.777, and 0.771), while wealth-dependence β nearly triples from 0.034 in the baseline to 0.091 in the full model, with most of that increase appearing once ICC is introduced (column 2). Cognitive noise falls dramatically, from 1.485 to 0.577 (a 61% reduction), confirming that simpler models conflate decision imprecision with preference heterogeneity. For girls, the pattern differs: r remains stable (from 0.546

to 0.553, 0.582, and 0.576) and β rises only modestly, from 0.080 in the baseline to 0.093 in the full model. The key difference is that girls' noise reduction is smaller (55% vs. 61%) and their ICC estimates are lower (0.383 vs. 0.428 for deliberation), suggesting that boys' deliberation noise is more trait-like.[20]

[20]The corresponding CRRA estimates are reported in the appendix: the baseline in Appendix K (Table K.1) and the estimates across all specifications in Appendix Tables O.3 and O.4.

Table 4: Building Toward the Full Model: Expo-Power Specification

	Model Specification			
Parameter	(1) Baseline	(2) +ICC	(3) +Guess	(4) Full
Panel A: Boys				
r in Expo-Power	0.767^{***}	0.760^{***}	0.777^{***}	0.771^{***}
	(0.055)	(0.034)	(0.029)	(0.033)
β in Expo-Power	0.034^{*}	0.082^{***}	0.099^{***}	0.091^{***}
	(0.018)	(0.007)	(0.007)	(0.009)
σ_{u^D}	1.485^{***}	1.046^{***}	0.519^{***}	0.577^{***}
	(0.128)	(0.072)	(0.034)	(0.046)
ICC^D	—	0.180^{***}	0.425^{***}	0.428^{***}
		(0.010)	(0.020)	(0.033)
ICC^G	—	—	0.632^{***}	0.574^{***}
			(0.023)	(0.025)
$\rho(\theta^D, \theta^G)$	—	—	-0.189^{***}	-0.150^{***}
			(0.044)	(0.034)
Pr(Guess)	—	—	0.126^{***}	0.144^{***}
			(0.012)	(0.014)
Panel B: Girls				
r in Expo-Power	0.546^{***}	0.553^{***}	0.582^{***}	0.576^{***}
	(0.025)	(0.024)	(0.023)	(0.025)
β in Expo-Power	0.080^{***}	0.086^{***}	0.095^{***}	0.093^{***}
	(0.004)	(0.004)	(0.005)	(0.005)
σ_{u^D}	2.013^{***}	1.720^{***}	0.850^{***}	0.908^{***}
	(0.163)	(0.134)	(0.062)	(0.082)
ICC^D	—	0.163^{***}	0.384^{***}	0.383^{***}
		(0.009)	(0.021)	(0.019)
ICC^G	—	—	0.575^{***}	0.531^{***}
			(0.023)	(0.023)
$\rho(\theta^D, \theta^G)$	—	—	0.056	0.108^{**}
			(0.084)	(0.045)
Pr(Guess)	—	—	0.164^{***}	0.187^{***}
			(0.014)	(0.012)

Notes: We compute predictions for each observation and average them over the sample. Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. All specifications use Expo-Power utility functions. The parameters are: r, the risk aversion curvature of the utility function; β, the wealth-dependence of relative risk aversion; σ_{u^D}, the standard deviation of transitory noise; ICC^D and ICC^G, the intra-class correlations of the persistent components of deliberation and guessing noise; $\rho(\theta^D, \theta^G)$, the correlation between those persistent components; and Pr(Guess), the mean probability of guessing. (1) *Baseline* includes only preference and noise parameters, without IQ or personality traits in any parameter. (2) *+ICC* includes the intra-class correlation structure for deliberation, but not guessing behavior or IQ and personality traits. (3) *+Guess* includes both the ICC structure and the guessing behavior, but not IQ and personality traits. (4) *Full* includes guessing behavior, ICC structures for both deliberation and guessing, guessing-deliberation correlation, and IQ and personality traits in each parameter. The corresponding CRRA estimates are reported in Appendix K (baseline) and Appendix Tables O.3 and O.4 (across specifications). $N = 1{,}115$ for boys and $N = 1{,}089$ for girls.

5.2.4 Values of Parameters Over the Sampling Distribution of IQ and Traits Across Specifications

Table 5 provides values of structural parameters averaged over the sampling distribution of IQ and traits across all model specifications for both genders. For each gender, we use the sample appropriate to it.

Panel A of Table 5 presents results for boys. Several patterns emerge. Adding IQ and personality traits absorbs part of the persistent component of guessing (ICC^G falls from 0.63 to 0.57) while leaving persistent deliberation differences essentially unchanged (ICC^D $\approx$ 0.43). The correlation between the persistent components of deliberation and guessing noise $Corr(\theta^D, \theta^G)$ is consistently negative (-0.15 to -0.19). Cognitive noise estimates are highly sensitive to model specification, with the full model yielding estimates 60–70% lower than baseline models. The comparisons for CRRA appear in Appendix Table O.3.

Panel B presents the corresponding results for girls. The patterns differ in important ways. Most strikingly, the correlation between persistent components of deliberation and guessing noise is positive for girls (+0.06 to +0.11), opposite to that of boys. Cognitive noise under Expo-Power is substantially higher for girls than boys (0.908 vs. 0.577), suggesting that girls' choices exhibit greater variability even after accounting for ICC structure and guessing. A comparison with CRRA appears in Appendix Table O.4. Appendix Table E.7 reports the estimated intra-class correlations in deliberation and guessing separately by gender.

Table 5: Values of Model Parameters over Sampling Distribution of IQ and Traits (Expo-Power)

	Parameter						
	$r(\boldsymbol{X})$	$\beta(\boldsymbol{X})$	$\sigma_{u^D}(\boldsymbol{X})$	ICC$^D(\boldsymbol{X})$	ICC$^G(\boldsymbol{X})$	Corr$(\theta^D,\theta^G)(\boldsymbol{X})$	Pr(Guess)$(\boldsymbol{X})$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Boys							
Expo-Power, with IQ and personality traits							
With ICC and Guessing	0.771***	0.091***	0.577***	0.428***	0.574***	-0.150^{***}	0.144***
	(0.033)	(0.009)	(0.046)	(0.033)	(0.025)	(0.034)	(0.014)
No ICC, with Guessing	0.793***	0.018	1.320***				0.090***
	(0.161)	(0.072)	(0.267)				(0.023)
With ICC, No Guessing	0.765***	0.074***	1.241***	0.190***			
	(0.097)	(0.027)	(0.226)	(0.010)			
No ICC, No Guessing	0.698***	0.043**	1.871***				
	(0.091)	(0.017)	(0.462)				
Expo-Power, without IQ or personality traits							
With ICC and Guessing	0.777***	0.099***	0.519***	0.425***	0.632***	-0.189^{***}	0.126***
	(0.029)	(0.007)	(0.034)	(0.020)	(0.023)	(0.044)	(0.012)
No ICC, with Guessing	0.711***	0.047***	1.227***				0.164***
	(0.041)	(0.009)	(0.098)				(0.018)
With ICC, No Guessing	0.760***	0.082***	1.046***	0.180***			
	(0.034)	(0.007)	(0.072)	(0.010)			
No ICC, No Guessing	0.767***	0.034*	1.485***				
	(0.055)	(0.018)	(0.128)				
Panel B: Girls							
Expo-Power, with IQ and personality traits							
With ICC and Guessing	0.576***	0.093***	0.908***	0.383***	0.531***	0.108**	0.187***
	(0.025)	(0.005)	(0.082)	(0.019)	(0.023)	(0.045)	(0.012)
No ICC, with Guessing	0.573***	0.083***	1.916***				0.073***
	(0.069)	(0.005)	(0.282)				(0.025)
With ICC, No Guessing	0.557***	0.087***	1.941***	0.166***			
	(0.030)	(0.004)	(0.171)	(0.009)			
No ICC, No Guessing	0.579***	0.080***	2.178***				
	(0.037)	(0.005)	(0.216)				
Expo-Power, without IQ or personality traits							
With ICC and Guessing	0.582***	0.095***	0.850***	0.384***	0.575***	0.056	0.164***
	(0.023)	(0.005)	(0.062)	(0.021)	(0.023)	(0.084)	(0.014)
No ICC, with Guessing	0.527***	0.078***	1.678***				0.143***
	(0.024)	(0.004)	(0.141)				(0.024)
With ICC, No Guessing	0.553***	0.086***	1.720***	0.163***			
	(0.024)	(0.004)	(0.134)	(0.009)			
No ICC, No Guessing	0.546***	0.080***	2.013***				
	(0.025)	(0.004)	(0.163)				

Notes: We compute predictions for each observation and average them over the sample. Each entry is the prediction over Sampling Distribution of IQ and traits from the Expo-Power model. Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Panel A reports boys ($N = 1{,}115$), Panel B girls ($N = 1{,}089$). The parameters are: r, the risk aversion curvature; β, the wealth-dependence of relative risk aversion; σ_{u^D}, transitory noise; ICC^D and ICC^G, the intra-class correlations of the persistent deliberation and guessing components; $Corr(\theta^D, \theta^G)$, their correlation; and Pr(Guess), the mean probability of guessing. For the CRRA specifications, see Appendix Tables O.3 and O.4.

5.3 How Do IQ and Personality Traits Affect Economic Preferences?

A central question of this paper is whether individual differences in IQ and personality systematically shape economic preferences and deliberation noise. We address this by examining covariate effects across all model parameters, using the full Expo-Power specification as our preferred model. We compare the specification for CRRA with that for Expo-Power in Appendix P.1.

5.3.1 Effects of IQ and Personality Traits on Relative Risk Aversion

Figure 4 presents the average marginal effects of IQ and personality traits on relative risk aversion under the full Expo-Power model. Because Expo-Power relative risk aversion $RRA(s + w) = r + \beta(s + w)^{1-r}$ varies with the stake-plus-wealth level, reporting it at a single stake would be arbitrary. We therefore evaluate $RRA(s + w)$ at $w = 1$ over a range of stakes $s \in \{1, 2, 5, 10, 20, 50, 100\}$[21]. The mean level of risk aversion rises with the stake for both genders, from 0.88 to 1.06 for boys and from 0.70 to 1.24 for girls. The gender difference therefore depends on the scale: at low stakes girls are less risk averse than boys, while at high stakes they are more risk averse, with the ordering reversing between $s = 20$ and $s = 50$. The gender gap in risk aversion is thus a function of the stakes at risk rather than a single number, a feature the risk aversion curvature r alone does not capture. Notice that for boys, the higher the stakes (and wealth), the weaker the effects of IQ and Conscientiousness, and the stronger the negative effects of Openness. For Agreeableness, Emotional Stability, and Extraversion, there is mild indication of an increasing effect of stakes on relative risk aversion. For girls, the pattern for IQ is similar to that of boys, but for Conscientiousness the pattern reverses. Both Agreeableness and Extraversion show decreasing effects with increasing stakes; there is a rising trend with

[21] Appendix G re-estimates all models across alternative reference wealth levels $w_0 \in \{0.2, 0.5, 1, 2, 5\}$ for both the Expo-Power and CRRA specifications. The level parameters—risk aversion curvature r and the transitory noise σ_{u^D}—shift mechanically with w_0, while the noise-structure parameters (ICC^D, ICC^G, ρ, and the guessing probability) and the gender pattern in risk attitudes are unchanged.

Openness—the opposite of the finding for boys.

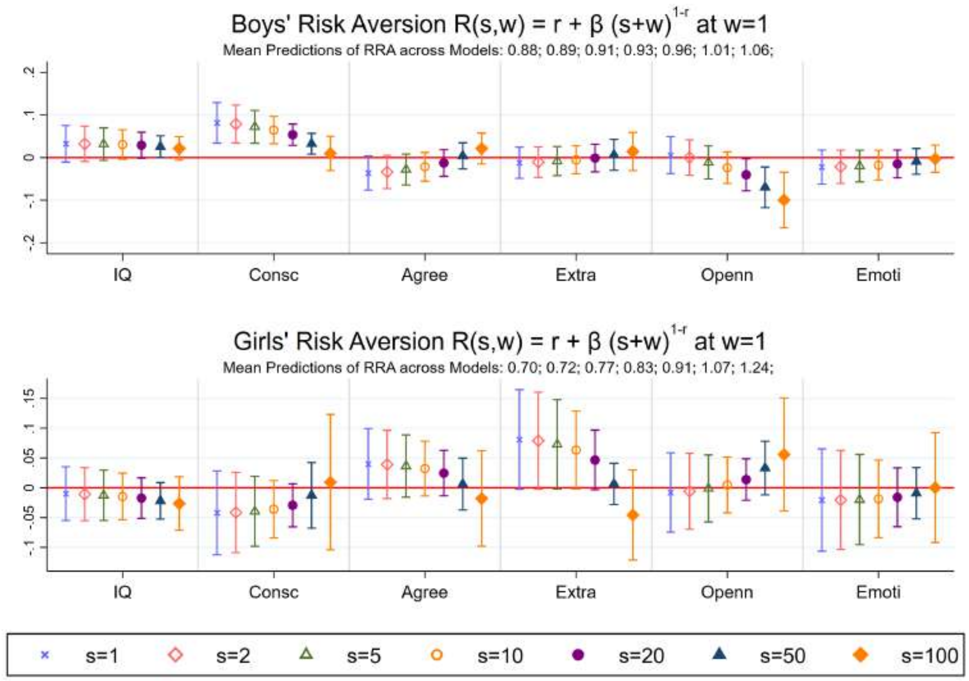


Figure 4: Effects of IQ and Personality Traits on Relative Risk Aversion $RRA(s + w)$ Across Stake Levels (s) Based on the Expo-Power Model, Evaluated by Mean of Gradients over the Sample

Notes: We compute the average of the gradients for each observation over the sample. Each marker is the average marginal effect of the indicated trait on relative risk aversion $RRA(s+w) = r + \beta(s + w)^{1-r}$ evaluated at $w = 1$, from the full Expo-Power model. Markers within each trait correspond to stakes $s = 1, 2, 5, 10, 20, 50, 100$. Top panel: boys. Bottom panel: girls. The subtitle of each panel reports the mean predicted $RRA(s+w)$ at each stake, evaluated at total wealth $s+w$ with $w = 1$ (so the stake $s = 1$ corresponds to $s+w = 2$). Error bars show 90% confidence intervals. N = 1,115 boys, N = 1,089 girls.

5.3.2 Effects of IQ and Personality Traits on r and β Parameters

Figure 5 decomposes the effects of IQ and personality traits on overall risk aversion into effects on risk aversion curvature r (top panels) and wealth-dependence β (bottom panels). This decomposition reveals whether traits affect the level of risk aversion, its wealth-dependence, or both.

For boys, Conscientiousness strongly increases r (coef. $= 0.074$, $p < 0.05$) with a modest positive effect on β (coef. $= 0.011$, $p < 0.10$). This indicates that conscientious boys exhibit higher risk aversion across all wealth levels, with slightly steeper wealth-dependence. Openness shows a significant negative effect on β (-0.028, $p < 0.05$) but no significant effect on r, suggesting that more open boys exhibit flatter risk aversion profiles. For girls, Extraversion increases both r (coef. $= 0.070$) and β (coef. $= 0.013$), though neither effect is statistically significant. The mean predictions reveal the source of gender differences: boys average $r = 0.77$ and $\beta = 0.09$, while girls average $r = 0.58$ and $\beta = 0.09$. Differences in risk parameters arise from differences in risk aversion curvature r, not wealth-dependence β.

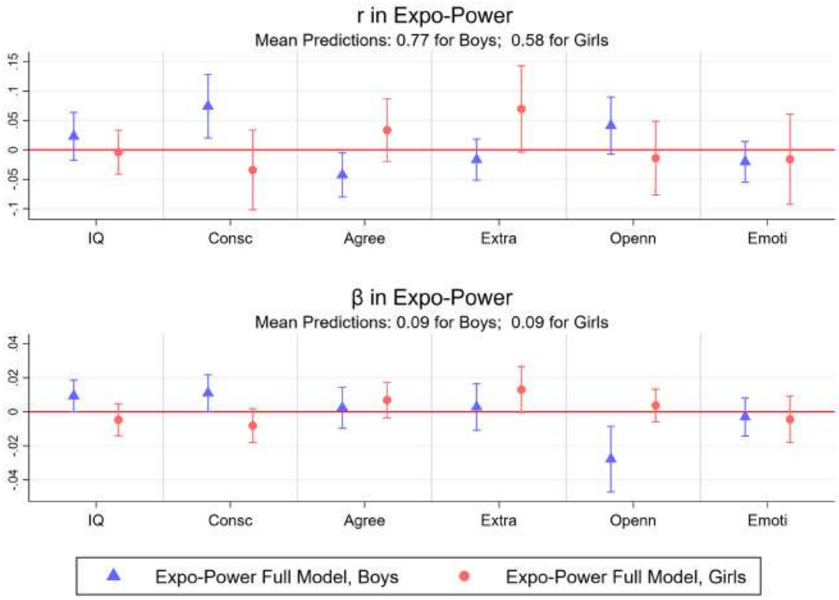


Figure 5: Decomposing Risk Aversion: Effects on r and β Parameters

Notes: We compute predictions for each observation and average them over the sample. Top panels show effects on risk aversion curvature r, bottom panels show effects on wealth-dependence β. Boys: Conscientiousness strongly increases r (coef. $= 0.074$, $p < 0.05$) with modest effect on β (coef. $=$ 0.011, $p < 0.10$). Girls: Extraversion increases both r (coef. $= 0.070$) and β (coef. $= 0.013$), neither significant. Mean predictions—Boys: $r = 0.77$, $\beta = 0.09$, Girls: $r = 0.58$, $\beta = 0.09$. Error bars show 90% confidence intervals.

5.3.3 Effects of IQ and Personality Traits on Deliberation Noise

Figure 6 presents the effects of IQ and personality traits on transitory deliberation noise σ_{u^D} (top panel) and the deliberation intra-class correlation ICC^D (bottom panel) under the full Expo-Power model. Conscientiousness and IQ reduce boys' deliberation noise, indicating more precise decisions. Mean predictions are $\sigma_{u^D} = 0.58$ for boys and 0.91 for girls, with $\text{ICC}^D = 0.43$ for boys and 0.38 for girls.

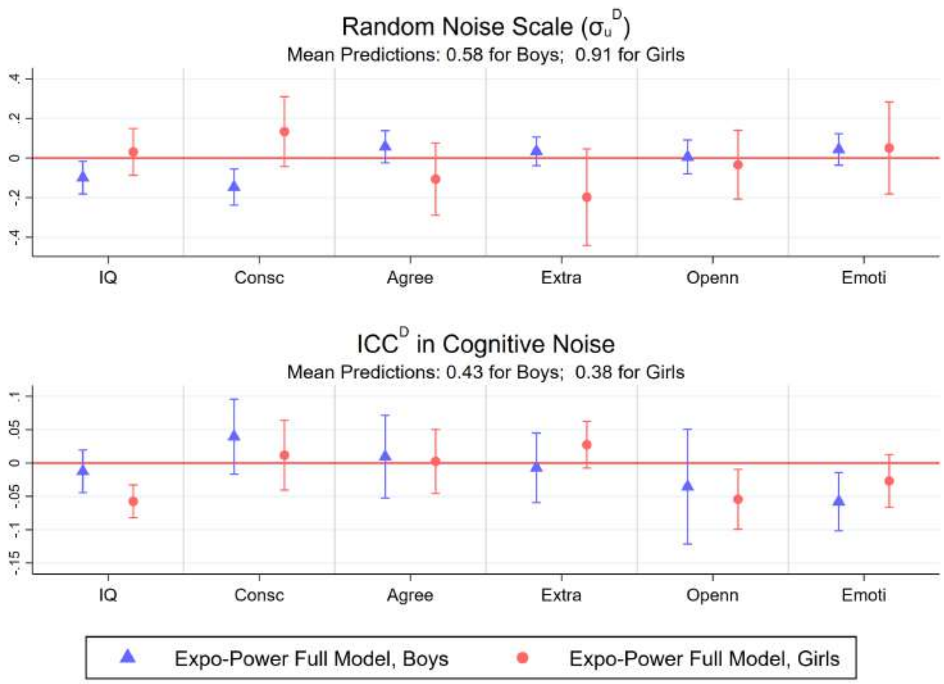


Figure 6: Trait Effects on Deliberation Noise: σ_{u^D} and ICC^D (Boys vs Girls)

Notes: We compute predictions for each observation and average them over the sample. Top panel shows effects on transitory noise σ_{u^D}, bottom panel on ICC^D, from the full Expo-Power model. Blue triangles: boys, red circles: girls. Error bars show 90% confidence intervals.

5.3.4 Effects of IQ and Personality Traits on Guessing Structure

Figure 7 presents the effects of IQ and personality traits on guessing behavior through ICC^G (top panels) and the correlation between persistent components of deliberation and guessing noise $Corr(\theta^D, \theta^G)$ (bottom panels).

For boys, Emotional Stability increases ICC^G (coef. $= 0.041$, $p < 0.10$), indicating that emotionally stable boys show more consistent guessing patterns. The baseline correlation is negative ($\rho = -0.15$). Extraversion reduces ICC^G (-0.061, $p < 0.10$).

For girls, the correlation structure differs fundamentally: the baseline correlation is positive ($\rho = +0.11$). IQ significantly strengthens this positive correlation (coef. $= 0.087$,

$p < 0.01$). These opposite correlation patterns suggest boys and girls employ fundamentally different cognitive strategies when facing difficult choices.

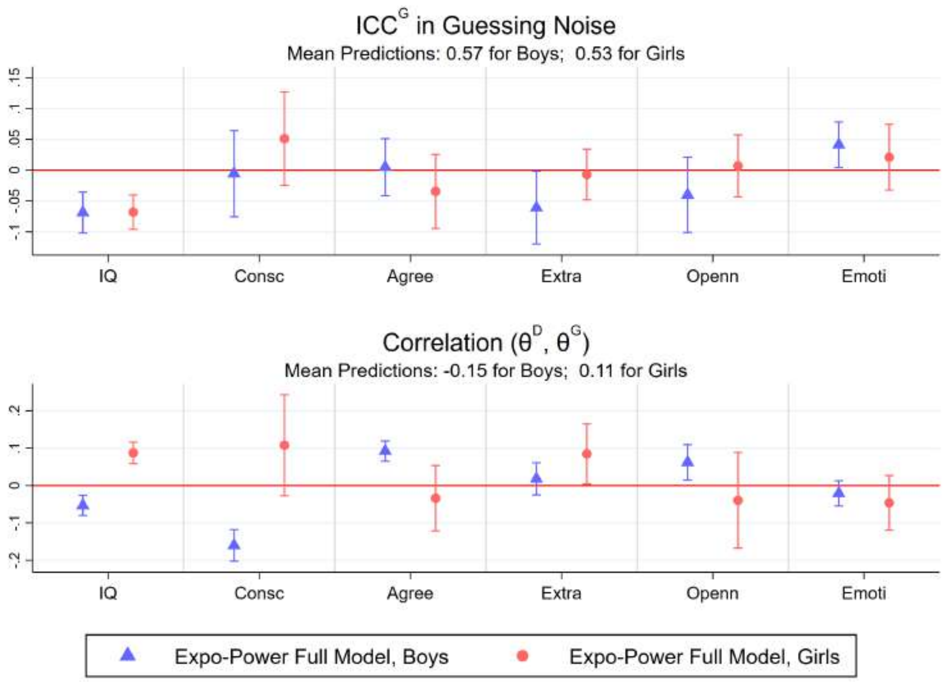


Figure 7: Trait Effects on Guessing Structure: ICC^G and Error Correlation (Boys vs Girls)

Notes: We compute predictions for each observation and average them over the sample. Top panel shows effects on ICC^G, bottom panel on the correlation $\rho = \text{Corr}(\theta^D, \theta^G)$, from the full Expo-Power model. Blue triangles: boys, red circles: girls. Error bars show 90% confidence intervals.

5.3.5 Effects of IQ and Personality Traits on Guessing

Figure 8 presents the mean effects of IQ and personality traits on overall guessing probability. IQ strongly reduces guessing for both genders (boys: -0.069, $p < 0.01$, girls: -0.096, $p < 0.01$), confirming that higher IQ promotes engagement with the decision task.

Beyond IQ, personality effects on guessing frequency are relatively weak. Openness reduces guessing for boys (-0.040, $p < 0.05$), while for girls only Conscientiousness (-0.036, $p < 0.10$) and Extraversion (-0.024, $p < 0.10$) weakly affect guessing probability. The weakness of these trait effects on overall frequency contrasts with significant effects on ICC^G and correlation (Figure 7), indicating that traits affect the consistency and structure of guessing rather than its frequency. Mean guessing probabilities are 14% for boys and 19% for girls, indicating that guessing behavior is a meaningful component of children's decision-making. This ordering is specific to the full model: without ICC structure the ranking reverses (for example, boys 16% versus girls 14% in the specification without ICC or traits), so girls' higher guessing frequency is a feature of our preferred specification rather than a universal pattern.

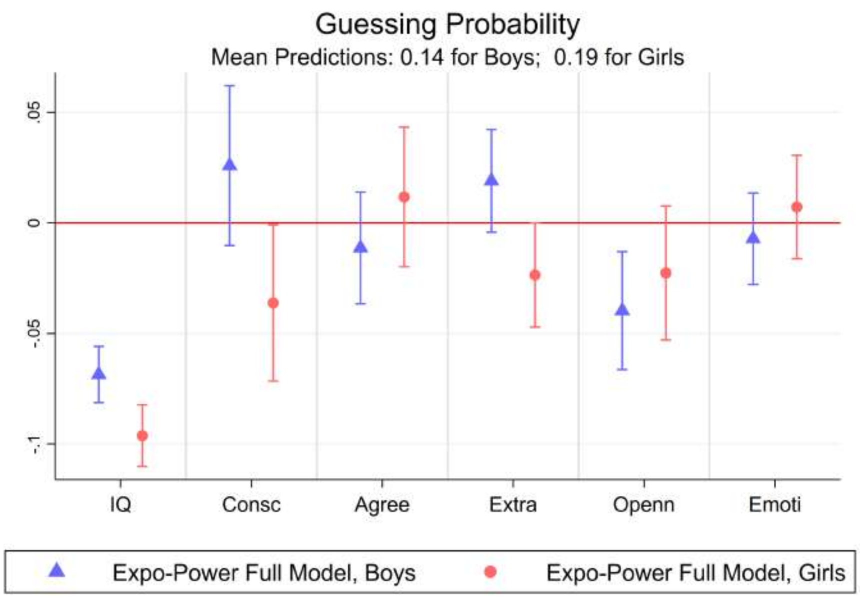


Figure 8: Trait Effects on Guessing Probability (Boys vs Girls)

Notes: We compute predictions for each observation and average them over the sample. Effects of IQ and traits on overall guessing probability, full Expo-Power model. Blue triangles: boys, red circles: girls. Mean guessing probabilities are 14% for boys and 19% for girls. Error bars show 90% confidence intervals.

5.3.6 Joint Significance Tests

Table 6 presents Wald tests for the joint significance of personality traits on each parameter for each gender for the Expo-Power model. Each entry tests the null hypothesis that the five personality traits jointly have no effect on the given parameter for that gender. This tests whether personality matters at all for a given parameter, not whether it operates differently across genders. We strongly reject the null that psychological traits do not enter the model (Column 8). Psychological traits jointly predict deliberation noise parameters (ICC, correlation, guessing) with more precision than risk preference parameters, particularly in the full model with ICC and guessing. Patterns differ by gender: for boys, traits significantly affect the ICCs in deliberation and guessing, the correlation between them, and the guessing probability, and also the risk aversion curvature r ($p = 0.028$), though this preference effect is smaller than the effects on the noise parameters; for girls, traits primarily affect the correlation structure and guessing probability. For comparison with the CRRA specification, see Appendix Table O.5.

Table 6: Joint Significance Tests: Personality Trait Effects on Model Parameters (Expo-Power)

	Parameter							
	$r(\boldsymbol{X})$ (1)	$\beta(\boldsymbol{X})$ (2)	$\sigma_{u^D}(\boldsymbol{X})$ (3)	$\text{ICC}^D(\boldsymbol{X})$ (4)	$\text{ICC}^G(\boldsymbol{X})$ (5)	$\text{Corr}(\theta^D, \theta^G)(\boldsymbol{X})$ (6)	$\Pr(\text{Guess})(\boldsymbol{X})$ (7)	All (8)
Boys, Expo-Power								
With ICC and Guessing	0.028	0.055	0.072	0.012	0.003	0.000	0.021	0.000
No ICC, with Guessing	0.032	0.705	0.903				0.008	0.000
With ICC, No Guessing	0.010	0.094	0.771	0.155				0.000
No ICC, No Guessing	0.010	0.016	0.569					0.000
Girls, Expo-Power								
With ICC and Guessing	0.496	0.053	0.247	0.329	0.607	0.000	0.006	0.000
No ICC, with Guessing	0.147	0.000	0.001				0.081	0.000
With ICC, No Guessing	0.000	0.000	0.000	0.175				0.000
No ICC, No Guessing	0.000	0.000	0.000					0.000

Notes: Each entry is a p-value for the joint null hypothesis that the five personality traits have no effect on the indicated parameter. Wald tests with 5 degrees of freedom. IQ enters the model but is not included in this joint test. Values below 0.05 indicate rejection at the 5% level. Empty cells indicate parameters not estimated in that specification. Column (8) tests whether all trait coefficients across all parameters are jointly zero. For the corresponding CRRA results, see Appendix Table O.5.

5.3.7 Source of Gender Differences

Table 7 addresses gender differences in the impact of traits. When boys and girls differ in a parameter, is this because they have different trait distributions (composition) or because traits operate differently by gender (structure)? A naïve comparison of predicted values at each gender's respective mean confounds these two sources.

Let each parameter q depend on the covariate vector through $q^g = F^q(\boldsymbol{X}^g, \boldsymbol{b}_g^q)$, where g indexes gender, $\boldsymbol{X}^g$ is the covariate vector (a constant plus IQ and personality traits) and $\boldsymbol{b}_g^q$ the coefficient vector for that parameter. We evaluate at the sample means and approximate the total change in q across genders by its first-order differential:

$$dq = \underbrace{\frac{\partial F^q}{\partial \boldsymbol{X}^g}\big(\boldsymbol{X}^g, \boldsymbol{b}_g^q\big)\, d\boldsymbol{X}^g}_{\text{Term 1}} \quad + \underbrace{\frac{\partial F^q}{\partial \boldsymbol{b}^q}\big(\boldsymbol{X}^g, \boldsymbol{b}_g^q\big)\, d\boldsymbol{b}_g^q}_{\text{Term 2}} \quad + \underbrace{\frac{\partial^2 F^q}{\partial \boldsymbol{b}^q\, \partial \boldsymbol{X}^g}\, d\boldsymbol{X}^g\, d\boldsymbol{b}_g^q}_{\text{Term 3}}. \tag{20}$$

Each term corresponds to one source in the threefold decomposition. Term 1 holds coefficients fixed and varies the mean covariates, giving the *composition* (endowment) source. Term 2 holds covariates fixed and varies the coefficients, giving the *coefficient* (structural) source. Term 3 is the *interaction*, the part depending jointly on both differences. All terms are evaluated at the boys' means $\bar{\boldsymbol{X}}_B$, with boys as the reference group; finite differences replace the differentials for binary covariates. Standard errors are obtained by the delta method.

For each parameter, Table 7 reports the four quantities in this decomposition—the total difference together with Terms 1, 2, and 3—as point estimates with significance stars.

Table 7: Sources of Girl–Boy Differences in Parameter Estimates
Evaluated at Mean Boy $\bar{\boldsymbol{X}}_B$ (Expo-Power)

	Parameters						
	$r(\boldsymbol{X})$	$\beta(\boldsymbol{X})$	$\sigma_{u^D}(\boldsymbol{X})$	$ICC^D(\boldsymbol{X})$	$ICC^G(\boldsymbol{X})$	$\text{Corr}(\theta^D, \theta^G)(\boldsymbol{X})$	$\text{Pr(Guess)}(\boldsymbol{X})$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Expo-Power, With ICC and Guessing							
Total	-0.196***	0.002	0.251***	-0.050	-0.046	0.267***	0.030
Term 1 (Composition)	0.025*	0.006*	-0.051***	0.009	0.006	-0.066***	0.006
Term 2 (Coefficient)	-0.183**	0.005	0.217	-0.047	-0.071	0.226	0.048
Term 3 (Interaction)	-0.039	-0.009**	0.085**	-0.012	0.019	0.107	-0.024**

Notes: For each parameter, we report the total girl–boy difference in predicted values together with Terms 1, 2, and 3 from Equation (20). The total is $F^q(\bar{\boldsymbol{X}}_G, \boldsymbol{b}^q_G) - F^q(\bar{\boldsymbol{X}}_B, \boldsymbol{b}^q_B)$, the difference in the predicted column parameter when each gender is evaluated at its own mean IQ and personality traits under its own coefficients. Term 1 (composition, endowments) is $\frac{\partial F^q}{\partial \boldsymbol{X}^g}(\bar{\boldsymbol{X}}_G - \bar{\boldsymbol{X}}_B)$, which holds coefficients at the boys' values and varies the mean covariates. Term 2 (coefficient, structure) is $\frac{\partial F^q}{\partial \boldsymbol{b}^q}(\boldsymbol{b}^q_G - \boldsymbol{b}^q_B)$, which holds covariates at the boys' means and varies the coefficients. Term 3 (interaction) is $\frac{\partial^2 F^q}{\partial \boldsymbol{b}^q \partial \boldsymbol{X}^g}(\bar{\boldsymbol{X}}_G - \bar{\boldsymbol{X}}_B)(\boldsymbol{b}^q_G - \boldsymbol{b}^q_B)$, the part that depends jointly on both differences. All derivatives are evaluated at the boys' means and coefficients $(\bar{\boldsymbol{X}}_B, \boldsymbol{b}^q_B)$. Linear decompositions of parameters involving nonlinear transformations are approximated using these differentials for continuous covariates and discrete differences for binary covariates, evaluated at the sample means. $\bar{\boldsymbol{X}}_B$ and $\boldsymbol{b}^q_B$ denote the mean covariate vector and coefficient vector for boys, and $\bar{\boldsymbol{X}}_G$ and $\boldsymbol{b}^q_G$ are the corresponding vectors for girls. Boys are the reference group. Standard errors are computed by the delta method. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. This table reports the full Expo-Power specification with ICC and guessing. For the full comparison across all four ICC and guessing combinations for both Expo-Power and CRRA, see Appendix Table O.6. Because these totals are evaluated at mean trait levels rather than averaged over the sample, they differ modestly from the average-over-sample (AME) parameter values in Table 5. For example, the total gender gap in σ_{u^D} is 0.251 here versus 0.331 from the sample-averaged predictions.

The decomposition reveals that gender differences in risk preferences reflect both compositional and structural sources, often operating in opposite directions. Compositional differences arise because boys and girls have different mean levels of traits (e.g., girls score higher on Conscientiousness); structural differences arise because the same trait maps differently to a parameter by gender (e.g., Conscientiousness raises r for boys but not girls). Term 1 isolates the endowment source, Term 2 the structural source, and Term 3 the part that depends jointly on both.

Under the full Expo-Power specification, girls have lower risk aversion curvature r than boys, a total difference of -0.196 ($p < 0.01$). The decomposition attributes almost all of this gap to Term 2, the coefficient component (-0.183, $p < 0.05$): the same traits map into r differently for girls. Term 1, the composition component, is small and works in the opposite direction (0.025, $p < 0.1$), so differences in trait endowments alone would narrow rather than widen the gap. For wealth-dependence β, the total difference is essentially zero and insignificant (0.002): boys and girls reach similar average β, with small composition and interaction terms offsetting one another. Girls also show noisier deliberation (σ_{u^D}: total difference $= 0.251$, $p < 0.01$) and a more positive correlation between the deliberation and guessing components (total difference $= 0.267$, $p < 0.01$); for both, Term 2 (the coefficient component) is the larger contributor while Term 1 (composition) pushes weakly the other way.[22]

5.3.8 Which Traits Matter and Through Which Channels?

Table 8 summarizes the statistically significant trait effects by gender and channel. IQ and personality traits matter for economic preferences, but they operate through different channels for boys and girls. For boys, traits affect both preferences and deliberation noise: Conscientiousness raises risk aversion curvature while Agreeableness lowers it, Openness lowers wealth-dependence, Conscientiousness sharpens deliberation, and IQ reduces guessing. Personality traits also shape the deliberation–guessing correlation most strongly of all—Conscientiousness and IQ push it further negative (Conscientiousness's -0.160 is the largest single trait effect in the model) while Agreeableness and Openness move it toward zero. For girls, traits operate almost entirely through the noise structure (ICCs and the deliberation–guessing correlation) and guessing rather than preferences—IQ shapes the persistence of deliberation and guessing, their correlation, and the guessing rate, while no trait reaches significance on risk preferences. This gender difference in which traits matter, confirmed by the significant coefficient components in Table 7, supports the conclusion that boys and girls exhibit distinct mappings from personality to economic behavior.

[22]The CRRA results appear in Appendix Table O.6.

Table 8: Summary of Significant Trait Effects

Gender	Trait	Channel	Effect
Boys	Conscientiousness	Risk aversion (r)	Increases (+0.074**)
	Agreeableness	Risk aversion (r)	Reduces (−0.042*)
	IQ	Wealth-dependence (β)	Increases (+0.009*)
	Conscientiousness	Wealth-dependence (β)	Increases (+0.011*)
	Openness	Wealth-dependence (β)	Reduces (−0.028**)
	IQ	Decision precision (σ_{u^D})	Improves (−0.099*)
	Conscientiousness	Decision precision (σ_{u^D})	Improves (−0.147***)
	Emotional Stability	Deliberation consistency (ICC^D)	Reduces (−0.058**)
	IQ	Guessing consistency (ICC^G)	Reduces (−0.069***)
	Emotional Stability	Guessing consistency (ICC^G)	Increases (+0.041*)
	Extraversion	Guessing consistency (ICC^G)	Reduces (−0.061*)
	IQ	Correlation (ρ)	More negative (−0.053***)
	Conscientiousness	Correlation (ρ)	More negative (−0.160***)
	Agreeableness	Correlation (ρ)	More positive (+0.092***)
	Openness	Correlation (ρ)	More positive (+0.062**)
	IQ	Guessing probability	Reduces (−0.069***)
	Openness	Guessing probability	Reduces (−0.040**)
Girls	IQ	Deliberation consistency (ICC^D)	Reduces (−0.057***)
	Openness	Deliberation consistency (ICC^D)	Reduces (−0.054**)
	IQ	Guessing consistency (ICC^G)	Reduces (−0.068***)
	IQ	Correlation (ρ)	More positive (+0.087***)
	Extraversion	Correlation (ρ)	More positive (+0.085*)
	IQ	Guessing probability	Reduces (−0.096***)
	Conscientiousness	Guessing probability	Reduces (−0.036*)
	Extraversion	Guessing probability	Reduces (−0.024*)

Notes: We compute predictions for each observation and average them over the sample. Significant effects are reported from the full Expo-Power model. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

5.4 Alternative Specifications of Guessing

We compare three specifications of equation (8): (1) guessing depends only on IQ and personality traits (the default setting in our full model), (2) guessing depends on deliberation noise ($\sigma^2_{u^D}$), with traits influencing guessing indirectly through their effect on deliberation noise, but with no direct effect, and (3) both channels operate jointly. In each of these models, the probability of guessing is specified using the same framework as in Equation (8), with differences captured in the function $\eta(\sigma^2_{u^D}(\boldsymbol{X}_i), \boldsymbol{X}_i)$ as described below:

Model 1 (standard full model, allowing IQ and traits to affect guessing directly):

$$\eta(\boldsymbol{X}_i) = \sum_{k=1}^{K} \eta_k \cdot \mathrm{X}_{ki}, \tag{21}$$

where η_k is the parameter associated with X_{ki}.

Model 2 (IQ and traits affect guessing indirectly, only through the deliberation noise σ_u^D), specifically:

$$\eta(\sigma^2_{u^D}(\boldsymbol{X}_i)) = \eta_0 + \eta_\sigma \cdot \sigma^D_{u,i}, \tag{22}$$

with $\sigma^2_{u^D}(\boldsymbol{X}_i)$ being the variance of the deliberation noise defined in Equation (19).

Model 3 (IQ and traits affect guessing through both channels in Models 1 and 2):

$$\eta(\sigma^2_{u^D}(\boldsymbol{X}_i), \boldsymbol{X}_i) = \eta_\sigma \cdot \sigma^D_{u,i} + \sum_{k=1}^{K} \eta_k \cdot \mathrm{X}_{ki}. \tag{23}$$

In all three models, we include ICC, guessing, and the effects of IQ and personality traits, consistent with our full baseline specification. The three specifications achieve very similar AIC values, with Model 1 marginally the best for boys and Model 3 marginally the best for girls. Table 9 reports the log-likelihood ratio test statistics comparing two pairs of models. The comparison of Model 3 with Model 1 addresses the central question, whether the effect of deliberation noise on guessing improves model fit beyond the direct trait effects. The comparison of Model 3 with Model 2 addresses the complementary question, whether the direct trait effects improve fit beyond the deliberation noise channel. Appendix Q presents detailed parameter estimates and coefficient plots for the three alternative specifications under the Expo-Power utility function. The models yield similar marginal effects of IQ and personality traits on most parameters, except for the guessing probability.

Table 9: Model Comparison: Alternative Guessing Specifications

Panel A: Boys	Model 1 Only Traits	Model 2 Only Cognitive Difficulty	Model 3 Joint
Loglikelihood	-32754.608	-32768.006	-32754.139
k (parameters)	49	44	50
AIC	65607.22	65624.01	65608.28
LR test: Model 3 vs Model 1 [$\chi^2(1)$]		0.94 ($p = 0.333$)	
LR test: Model 3 vs Model 2 [$\chi^2(6)$]		27.73 ($p < 0.001$)	

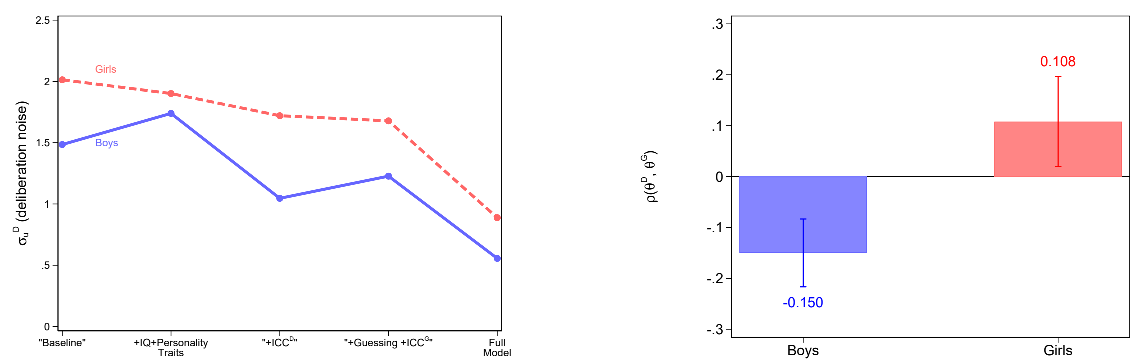
Panel B: Girls	Model 1 Only Traits	Model 2 Only Cognitive Difficulty	Model 3 Joint
Loglikelihood	-33232.961	-33239.669	-33231.398
k (parameters)	49	44	50
AIC	66563.92	66567.34	66562.80
LR test: Model 3 vs Model 1 [$\chi^2(1)$]		3.13 ($p = 0.077$)	
LR test: Model 3 vs Model 2 [$\chi^2(6)$]		16.54 ($p = 0.011$)	

Notes: All models incorporate the ICC structure and guessing behavior and are based on the Expo-Power utility function. Model 1 allows traits to affect guessing directly (equation (21)). Model 2 specifies that deliberation noise directly influences guessing, with traits affecting guessing only indirectly through their impact on deliberation noise (equation (22)). Model 3 includes both channels (equation (23)). Model 3 nests Models 1 and 2, so the likelihood ratio tests have 1 and 6 degrees of freedom. Models 1 and 2 are non-nested and are compared by AIC, where lower AIC indicates better fit.

In results displayed in Appendix Q, for girls, the marginal effect of deliberation noise σ_{u^D} on guessing probability is sizable, though only marginally significant, when traits are restricted to operate through noise variances alone (Model 2): under Expo-Power, the coefficient on σ_{u^D} in the guessing index is $\eta_\sigma = 2.857$ (SE 1.622, $p < 0.10$). When direct trait effects are added (Model 3), the coefficient weakens to -0.658 (SE 1.051). For boys, the deliberation noise channel is weak throughout: Model 2 yields a positive but insignificant coefficient ($\eta_\sigma = 0.296$, SE 0.276), and Model 3 a negative, insignificant one (-0.410, SE 0.983). The point estimates suggest that girls' guessing responds to imprecise deliberation signals while boys' does not, though the likelihood ratio test does not reject the null of no effect of cognitive noise variance on guessing. The finding that direct trait effects absorb the noise channel for both genders indicates that cognitive and personality characteristics influence guessing through multiple pathways, not solely through their impact on deliberation precision.

5.5 The Gender Gap

Figure 9 consolidates our findings on decision noise. Our analysis challenges conventional wisdom on the gender gap in risk aversion. Standard CRRA without accounting for deliberation noise replicates the traditional pattern (girls' single r 12% higher). However, the full CRRA model that adds correlated shocks, guessing, and traits eliminates the gap entirely (both $RRA(s+w) \approx 0.97$). Moreover, the best-fitting Expo-Power model reveals a reversal: over a range of values, girls are relatively less risk averse over small stakes and low wealth. They become more risk averse as the scale increases (Figures 2 and 3). The ICC structure reveals that approximately 40% of deliberation noise variance arises from persistent individual differences rather than random shocks. Boys and girls exhibit opposite correlation patterns in the persistent components of deliberation and guessing noise ($\rho = -0.15$ vs. $\rho = +0.11$).



(A) Deliberation Noise Reduction

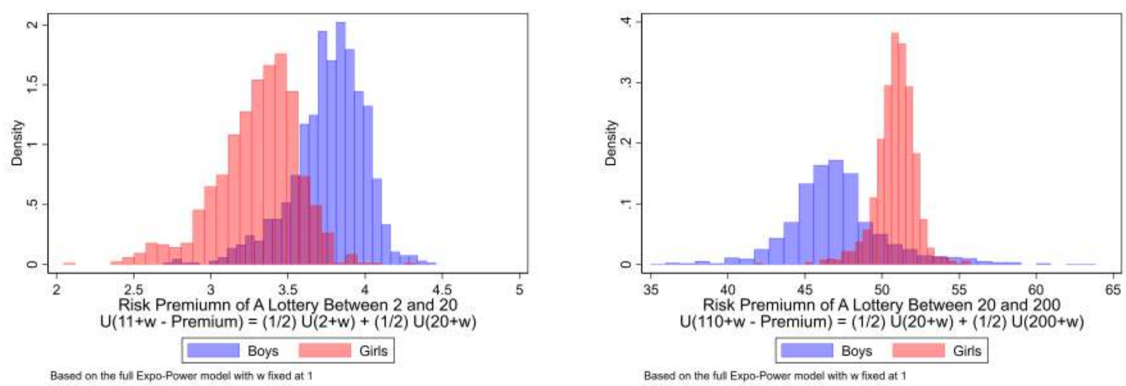
(B) Deliberation-Guessing Correlation

Figure 9: Deliberation Noise and Decision Processes

Notes: We compute predictions for each observation and average them over the sample. Panel A shows how estimated deliberation noise σ_{u^D} falls as model components are added across specifications, for boys and girls. The component interactions underlying these reductions are documented in Appendix R. Panel B shows the estimated correlation between the persistent component of errors in deliberation and guessing for boys and girls. Error bars represent 90% confidence intervals.

5.6 Gender Heterogeneity in Risk Preferences

Figure 10 shows histograms of willingness to pay to avoid risk for girls and boys at different levels of wealth and stakes, implied by the full Expo-Power estimates, for two of the 50–50 lotteries used in the survey, panel (a) for a lottery over 2 and 20 tokens and panel (b) for a lottery over 20 and 200 tokens. For each child the risk premium is the amount by which the certainty equivalent falls below the expected value of the lottery, computed from individual-specific covariates and estimated parameters. Panel (a) shows that at lower levels, girls tend to be less risk averse. Panel (b) shows that at higher levels, this is reversed. No single risk aversion parameter characterizes gender differences. The elicitation protocol for these lotteries is described in Appendix B.



(a) Lottery over 2 and 20 tokens

(b) Lottery over 20 and 200 tokens

Figure 10: Distribution of Implied Risk Premiums by Gender

Notes: Risk premiums implied by the full Expo-Power specification and individual traits, for two 50–50 survey lotteries: (a) over 2 and 20 tokens and (b) over 20 and 200 tokens. Boys (blue), girls (red). A higher premium indicates greater risk aversion. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

5.7 Robustness

Our main findings are robust to alternative specifications of socioeconomic variables, reference wealth levels, and reference wealth heterogeneity. We summarize each robustness check below. Full results appear in Appendices F–H. An important body of the literature in behavioral economics (see e.g., Malmendier and Nagel 2011; Malmendier and Wachter 2024) documents the role of background and personal experiences in shaping preferences. We find that background affects psychological traits, but conditioning on those traits there is no additional role for background variables. When added to models that already include traits, they have no effect on any component of our models. The background coefficients in model 3 are statistically indistinguishable from zero across all parameters, while trait effects remain stable. This indicates that family background influences risk attitudes and deliberation noise indirectly—by shaping IQ and personality that affect noise and guessing—rather than through direct channels. The full analysis appears in Appendix F.

A related concern is whether children from different socioeconomic backgrounds develop systematically different reference points based on their lived experiences. Appendix H examines this by comparing two specifications: model 1 fixes reference wealth $w = 1$ for all children, while model 2 estimates w as a flexible function of background variables including urban residence, parental education, sibling presence, and left-behind status.

An interesting finding emerges from the estimated reference wealth function itself. For boys, predicted reference wealth over the sampling distribution of background variables is $w = 2.13$, suggesting boys may integrate the experimental stakes within a broader consumption choice. For girls, predicted reference wealth is considerably lower ($w = 0.74$), indicating that girls may weight the experimental stakes more heavily relative to background consumption. This gender difference in reference point formation—boys integrating stakes with background wealth, girls evaluating stakes more narrowly—could contribute to observed differences in risk-taking behavior, though the mechanism warrants further investigation (see Chávez et al. (2026)).

The stability of structural parameters across reference wealth specifications reinforces our main conclusions. Whether reference wealth is fixed or allowed to vary with socioeconomic background, the key findings persist: ICC structure captures approximately 40%

of deliberation variance and 55% of guessing variance, boys and girls show opposite-signed correlations between deliberation and guessing, and the Expo-Power specification reveals that boys are more risk averse than girls at small stakes and low wealth once deliberation noise is properly modeled.

6 Choice, Mistakes, and Utility Losses

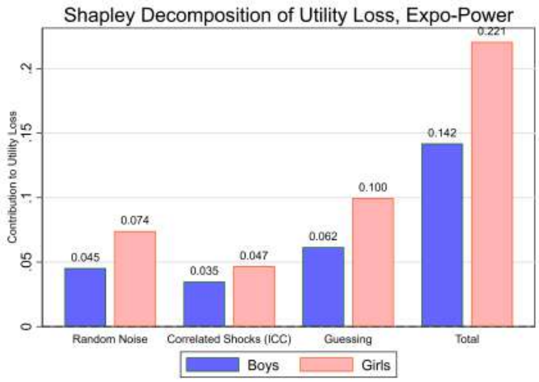
We next quantify the relative importance of deliberation noise, guessing, and preferences for explaining observed choices in the fully developed model. We assume that our estimates reveal true parameters. We present a sequential decomposition in Appendix S. In the text, we present an order-invariant Shapley decomposition.

6.1 Shapley Decomposition

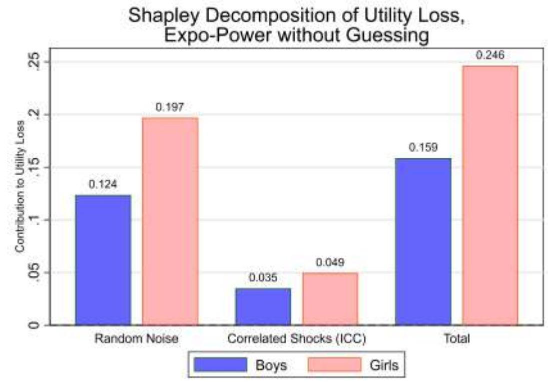
Figure 11 presents the Shapley value decomposition, which computes each component's average marginal contribution across all possible orderings, addressing the path dependence problem that arises from the use of any particular sequential decomposition. We report results for the Expo-Power model. Results for CRRA are virtually identical.[23]

The total choice-mistake rate is about 30%. Of this, random transitory noise (u_{ij}^D) accounts for approximately 11–13 percentage points across both genders and specifications, persistent individual noise (θ_i^D) for roughly 8–10, and guessing behavior for approximately 9.

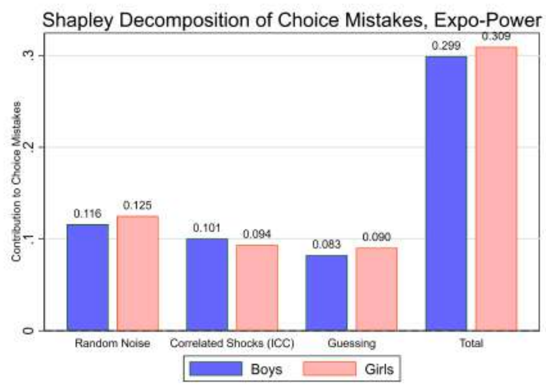
[23] The two specifications yield virtually identical decompositions of *choice mistakes*: total mistake rates are 0.299 (boys) and 0.309 (girls) under Expo-Power versus 0.300 and 0.310 under CRRA. For *utility loss*, however, they diverge: the guessing contribution and total utility loss are larger under Expo-Power (0.062 and 0.100, with totals 0.142 and 0.221 for boys and girls) than under CRRA (0.051 and 0.054, with totals 0.117 and 0.124), with the gap concentrated among girls. See Appendix S.



(a) Utility Loss



(b) Utility Loss in a Model without Guessing



(c) Choice Mistakes

Figure 11: Shapley Decomposition of Choice Mistakes and Utility Loss (Expo-Power)

Notes: Shapley value decomposition for utility loss (panel a), for utility loss in a model estimated without guessing (panel b), and for choice mistakes (panel c). Blue bars: boys, red bars: girls. Each component's contribution is its average marginal effect across all possible orderings. "Random Noise" refers to transitory noise u_{ij}^D, "Persistent Noise" to individual noise θ_i^D, and "Guessing" to guessing behavior. "Total" shows the sum. In panel (b), the model omits guessing, and its contribution is absorbed by the noise components.

All three error sources make substantial and roughly comparable contributions. Transitory noise is the largest single contributor, but persistent noise and guessing each account for a substantial amount of variation. This contradicts models that attribute all choice inconsistency to pure randomness. Guessing is a novel and important feature of our analysis, especially for girls. When we drop guessing, it is unsurprising that the contribution of deliberation noise increases substantially for mistakes and utility losses. Gender differences are modest. Girls have higher contributions from transitory noise (12.5%) than boys (11.6%), while persistent noise and guessing show similar contributions. In terms of utility loss rather than mistake counts, guessing becomes the largest single contributor (0.062 for boys and 0.100 for girls), exceeding both random and persistent noise, and the gender gap widens (total utility loss of 0.142 for boys versus 0.221 for girls).

6.2 Implications for Preference Measurement

These decompositions carry important methodological implications. Traditional approaches that attribute all choice inconsistency to independent shocks to preferences systematically overestimate the role of noise and underestimate the contributions of persistent individual differences and guessing behavior identified in this paper. Our findings suggest approximately 30% of choices are mistakes, with the largest share coming from transitory noise, followed by guessing and correlated noise, which might be interpreted as arising from unobserved personality traits.

This evidence has practical consequences. First, researchers should account for persistent heterogeneity in deliberation noise when estimating preference parameters and when using them in subsequent investigations. Second, guessing is an important source of error. However, guessing probabilities fundamentally depend on traits and are not a simple binomial probability as assumed in most previous work. Third, the substantial role of uncorrelated noise suggests that within-subject repeated measures can improve precision by averaging over task-specific noise, though persistent components remain. They might be fruitfully modeled by fixed effects.

The robustness of these decompositions across both CRRA and Expo-Power specifications demonstrates that the underlying decision errors are not artifacts of functional form assumptions. Whether risk aversion is constant or wealth-dependent, the same three sources account for observed mistakes in roughly the same proportions.

7 Summary and Conclusions

This paper addresses a fundamental challenge in preference measurement: when deliberation noise varies systematically across individuals, standard approaches confound true preference heterogeneity with differences in the precision with which those preferences are expressed. We address this problem in the context of recovering the true risk preferences of Chinese children.

We establish seven main results:

(1) The Expo-Power function describes the risk-taking behavior of young children better than the widely used CRRA model, especially for girls but also for boys in our fully specified model.

(2) Psychological traits affect preferences primarily through their effects on deliberation noise and guessing. IQ and personality traits operate through multiple channels that vary by gender. For boys, traits primarily affect deliberation precision, guessing, and error correlation parameters but also affect preference parameters. For girls, traits operate almost entirely through the noise structure (ICCs and the deliberation-guessing correlation) and guessing, not through preferences.

(3) Background variables do not directly affect preferences. They affect the psychological traits, which in turn primarily affect deliberation and guessing.

(4) Shocks to preferences featured in the literature are strongly correlated across choices. This is consistent with the notion that 'noise' featured in the recent literature (see e.g., Woodford (2020)) is substantially due to missing personality traits that are not captured by the measures available to us.

(5) Guessing behavior is an important source of decision errors, distinct from deliberation noise. Moreover, deliberation noise itself is not purely transitory. We decompose total decision variance into persistent and transitory components, finding that approximately

40% of deliberation error reflects stable individual differences in deliberation noise. Baseline models that treat all inconsistency as transitory noise severely overestimate cognitive imprecision. Once we separate these components, girls have higher deliberation noise than boys even after controlling for persistent noise.

(6) The best-fitting Expo-Power model reveals that gender differences in risk aversion depend on stakes and wealth. Boys are more risk averse in small-stakes bets. Girls are more risk averse in large-stakes bets.

(7) Components interact synergistically rather than additively. The full model improves fit substantially beyond the sum of individual components, confirming that proper identification of preferences requires simultaneous treatment of preferences, deliberation and guessing. "Measurement error" is not a simple mean-zero, independent, classical econometric error. It arises from guessing and deliberation noise, which in turn depend on psychological and background variables.

These findings likely extend beyond childhood preference measurement. Analysis of populations exhibiting systematic variation in cognitive constraints or task engagement requires frameworks that jointly model preferences and decision processes. Elderly populations facing cognitive decline, clinical groups with attention disorders, and survey respondents under time pressure all present similar challenges where deliberation noise varies systematically.

The structural approach we develop provides a template for settings where deliberation noise varies systematically. Our quantitative findings establish that accounting for decision processes is not a statistical refinement but a prerequisite for accurate preference recovery. When heterogeneity extends beyond tastes to the precision and consistency with which preferences are expressed, measuring one without the other distorts conclusions about both.[24]

[24]Embedded in the models we fit is a 'wealth parameter'. What this means for children is not obvious. This finding is explained in Chávez et al. (2026), who develop an approach that interprets the substitution of stakes and family background in a more flexible way. They consider more general models of child risk taking and explore a wider class of models that include loss aversion.

References

Alan, Sule, and Seda Ertac. 2018. "Fostering Patience in the Classroom: Results from Randomized Educational Intervention." *Journal of Political Economy* 126 (5): 1865–1911.

Almlund, Mathilde, Angela L. Duckworth, James J. Heckman, and Tim Kautz. 2011. "Personality Psychology and Economics." Chap. 1 in *Handbook of the Economics of Education,* edited by Eric A. Hanushek, Stephen Machin, and Ludger Wößmann, 4:1–181. Amsterdam: Elsevier B. V. https://doi.org/10.1016/B978-0-444-53444-6.00001-8.

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

Andersson, Ola, Hakan J. Holm, Jean-Robert Tyran, and Erik Wengström. 2016. "Risk aversion relates to cognitive ability: Preferences or noise?" *Journal of the European Economic Association* 14 (5): 1129–1154.

———. 2020. "Robust Inference in Risk Elicitation Tasks." *Journal of Risk and Uncertainty* 61 (3): 195–209. https://doi.org/10.1007/s11166-020-09341-6.

Arrow, Kenneth J. 1965. *Aspects of the Theory of Risk-Bearing.* Helsinki: Yrjö Jahnsson Foundation.

Barseghyan, Levon, Francesca Molinari, Ted O'Donoghue, and Joshua C. Teitelbaum. 2018. "Estimating Risk Preferences in the Field." *Journal of Economic Literature* 56 (2): 501–564. https://doi.org/10.1257/jel.20161148.

Belzil, Christian, and Tomáš Jagelka. 2024. "Separating Preferences from Endogenous Effort and Cognitive Noise in Observed Decisions." CREST Working Papers Series No. 2024-13.

Borghans, Lex, Angela L. Duckworth, James J. Heckman, and Bas ter Weel. 2008. "The Economics and Psychology of Personality Traits." *Journal of Human Resources* 43, no. 4 (Fall): 972–1059. https://doi.org/10.3368/jhr.43.4.972.

Caplin, Andrew. 2025. *An Introduction to Cognitive Economics: The Science of Mistakes.* Cham: Palgrave Macmillan.

Chávez, Cesar, Shuaizhang Feng, James J. Heckman, and Zhe Yang. 2026. "Alternative Models for the Risk Preferences of Young Children." Department of Economics, University of Chicago.

Chernoff, Herman. 1954. "On the Distribution of the Likelihood Ratio." *The Annals of Mathematical Statistics* 25 (3): 573–578. https://doi.org/10.1214/aoms/1177728725.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. 2011. "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences." *Journal of the European Economic Association* 9 (3): 522–550.

Enke, Benjamin, Uri Gneezy, Brian Hall, David C. Martin, Vadim Nelidov, Theo Offerman, and Jeroen van de Ven. 2023. "Cognitive Biases: Mistakes or Missing Stakes?" *The Review of Economics and Statistics* 105 (4): 818–832. https://doi.org/10.1162/rest_a_01093.

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.

Feng, Shuaizhang, Yujie Han, James J. Heckman, and Tim Kautz. 2022. "Comparing the reliability and predictive power of child, teacher, and guardian reports of noncognitive skills." *Proceedings of the National Academy of Sciences* 119 (6): e2113992119. https://doi.org/10.1073/pnas.2113992119.

Fomby, Paula, and Andrew J Cherlin. 2007. "Family instability and child well-being." *American Sociological Review* 72 (2): 181–204.

Harbaugh, William T., Kate Krause, and Lise Vesterlund. 2002. "Risk Attitudes of Children and Adults: Choices Over Small and Large Probability Gains and Losses." *Experimental Economics* 5 (1): 53–84.

Harless, David W., and Colin F. Camerer. 1994. "The Predictive Utility of Generalized Expected Utility Theories." *Econometrica* 62 (6): 1251–1289.

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

Heckman, James J., Tomáš Jagelka, and Timothy D. Kautz. 2021. "Some Contributions of Economics to the Study of Personality." In *Handbook of Personality: Theory and Research,* 4th, edited by Oliver P. John and Richard W. Robins, 853–892. New York: Guilford Press.

Heckman, James J., Rodrigo Pinto, and Peter A. Savelyev. 2013. "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review* 103, no. 6 (October): 2052–2086. https://doi.org/10.1257/aer.103.6.2052.

Hey, John D., and Chris Orme. 1994. "Investigating Generalizations of Expected Utility Theory Using Experimental Data." *Econometrica* 62 (6): 1291–1326.

Hill, Martha S, Wei-Jun J Yeung, and Greg J Duncan. 2001. "Childhood family structure and young adult behaviors." *Journal of Population Economics* 14 (2): 271–299.

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

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

Kautz, Tim, and Wladimir Zanoni. 2014. "Measuring and Fostering Non-Cognitive Skills in Adolescence: Evidence from Chicago Public Schools and the OneGoal Program." Unpublished manuscript, University of Chicago, Department of Economics.

Loewenstein, George. 2000. "Emotions in Economic Theory and Economic Behavior." *American Economic Review* 90 (2): 426–432.

Lord, Frederic M., and Melvin R. Novick. 1968. *Statistical Theories of Mental Test Scores.* Reading, MA: Addison-Wesley.

Malmendier, Ulrike, and Stefan Nagel. 2011. "Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?" *The Quarterly Journal of Economics* 126 (1): 373–416. https://doi.org/10.1093/qje/qjq004.

Malmendier, Ulrike, and Jessica A. Wachter. 2024. "Memory of Past Experiences and Economic Decisions." In *The Oxford Handbook of Human Memory,* edited by Michael J. Kahana and Anthony D. Wagner. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190917982.013.78.

Pratt, John W. 1964. "Risk Aversion in the Small and in the Large." *Econometrica* 32 (1-2): 122–136. https://doi.org/10.2307/1913738.

Rabin, Matthew. 1998. "Psychology and Economics." *Journal of Economic Literature* 36 (1): 11–46.

Saha, Atanu. 1993. "Expo-Power Utility: A 'Flexible' Form for Absolute and Relative Risk Aversion." *American Journal of Agricultural Economics* 75 (4): 905–913. https://doi.org/10.2307/1243978.

Self, Steven G., and Kung-Yee Liang. 1987. "Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions." *Journal of the American Statistical Association* 82 (398): 605–610. https://doi.org/10.1080/01621459.1987.10478472.

Sims, Christopher A. 2003. "Implications of Rational Inattention." *Journal of Monetary Economics* 50 (3): 665–690.

Sutter, Matthias, Claudia Zoller, and Daniela Glätzle-Rützler. 2019. "Economic behavior of children and adolescents–A first survey of experimental economics results." *European Economic Review* 111:98–121.

Thurstone, Louis L. 1927. "A law of comparative judgment." *Psychological Review* 34 (4): 273–286. https://doi.org/10.1037/h0070288.

Von Gaudecker, Hans-Martin, Arthur Van Soest, and Erik Wengström. 2011. "Heterogeneity in risky choice behavior in a broad population." *American Economic Review* 101 (2): 664–694.

Woodford, Michael. 2020. "Modeling Imprecision in Perception, Valuation, and Choice." *Annual Review of Economics* 12 (1): 579–601.

Xie, Danyang. 2000. "Power Risk Aversion Utility Functions." Working Paper version: CEMA Working Paper No. 22 (Oct 1999), *Annals of Economics and Finance* 1, no. 2 (November): 265–282. https://aeconf.com/Articles/Nov2000/aef010202.pdf.

Online Appendix

of *Relating Cognitive Skills and Personality Traits to Economic Preferences: Decision-Making under Cognitive Shocks and Guessing*

A Log-Likelihood and Identification

This appendix first defines the error structure and the parameter vector, assembles the log-likelihood, gives a constructive argument linking each parameter to the features of the data that identify it, and sketches a proof of identification. The model is a mixture of a deliberation regime and a guessing regime, integrated over two latent random effects.

A.1 Error Structure

The model specifies the probability of choice for each item, generated by latent random variables that govern deliberation and guessing. Stacking the J items for person i, the deliberation and guessing shocks are

$$\varepsilon^D = \begin{pmatrix} \varepsilon^D_{i1} \\ \varepsilon^D_{i2} \\ \vdots \\ \varepsilon^D_{iJ} \end{pmatrix} = \theta^D_i \boldsymbol{\iota} \quad + \underbrace{\begin{pmatrix} u^D_{i1} \\ u^D_{i2} \\ \vdots \\ u^D_{iJ} \end{pmatrix}}_{\boldsymbol{U}^D}, \qquad \varepsilon^G = \begin{pmatrix} \varepsilon^G_{i1} \\ \varepsilon^G_{i2} \\ \vdots \\ \varepsilon^G_{iJ} \end{pmatrix} = \theta^G_i \boldsymbol{\iota} \quad + \underbrace{\begin{pmatrix} u^G_{i1} \\ u^G_{i2} \\ \vdots \\ u^G_{iJ} \end{pmatrix}}_{\boldsymbol{U}^G}, \tag{A.1}$$

where $\boldsymbol{\iota}$ is a $J \times 1$ vector of ones, with $J = 56$ binary items nested in six price lists, and θ^D_i and θ^G_i are the person-level persistent components. $\boldsymbol{U}^D$ and $\boldsymbol{U}^G$ collect the transitory item-level shocks.

A.1.1 Guessing in the Likelihood

The guessing process is specified in Section 2 of the main text; here we restate only the objects required to form the likelihood, in the stacked notation introduced above. For person i the guessing shock on item j is $\varepsilon^G_{ij} = \omega_j\, \theta^G_i + u^G_{ij}$, with the exchangeability normalization $\omega_j = 1$, so that the full J-vector of guessing shocks is $\boldsymbol{\varepsilon}^G_i = \theta^G_i\, \boldsymbol{\iota} + \boldsymbol{U}^G_i$, paralleling the deliberation shocks $\boldsymbol{\varepsilon}_i^D = \theta_i^D \boldsymbol{\iota} + \boldsymbol{U}_i^D$. The guessing indicator on item j is

$$G_{ij} = \mathbb{1}\Big(\eta(\sigma^2_{u_i^D}, \boldsymbol{X}_i) + \omega_j\, \theta_i^G + u_{ij}^G > 0\Big), \tag{A.2}$$

with $\eta(\cdot)$ as defined in the main text. The persistent components enter the covariance structure exactly as for deliberation:

$$\mathrm{Var}(\boldsymbol{\varepsilon}_i^G) = \sigma^2_{\theta^G}\, \boldsymbol{\iota}\boldsymbol{\iota}' + \sigma^2_{u^G}\, \boldsymbol{I}_J, \qquad ICC^G = \frac{\sigma^2_{\theta^G}}{\sigma^2_{\theta^G} + \sigma^2_{u^G}}, \tag{A.3}$$

and the deliberation and guessing persistent components are jointly normal with correlation $\rho = \mathrm{Corr}(\theta_i^D, \theta_i^G)$. The shocks u_{ij}^D and u_{ij}^G are independent across (i, j) and across persons.

A.2 The Log-Likelihood

The log-likelihood function is given by the integral of the joint probability over the latent random effects (θ^D, θ^G):

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\Omega}) = \sum_{i=1}^{I} \ln \int\!\!\int \prod_{j=1}^{J} \Bigg\{ & \Big[\Pr(Y_{ij} = B \mid \theta^D, \boldsymbol{X}_i, G_{ij} = 0)\, \Pr(G_{ij} = 0 \mid \boldsymbol{X}_i, \theta^G, \eta_{ij}) \\
& + \tfrac{1}{2} \Pr(G_{ij} = 1 \mid \boldsymbol{X}_i, \theta^G, \eta_{ij})\Big]^{y_{ij}} \\
& \times \Big[\Pr(Y_{ij} = A \mid \theta^D, \boldsymbol{X}_i, G_{ij} = 0)\, \Pr(G_{ij} = 0 \mid \boldsymbol{X}_i, \theta^G, \eta_{ij}) \\
& + \tfrac{1}{2} \Pr(G_{ij} = 1 \mid \boldsymbol{X}_i, \theta^G, \eta_{ij})\Big]^{1 - y_{ij}} \Bigg\} \times\, dF(\theta^D, \theta^G),
\end{aligned} \tag{A.4}$$

where $\boldsymbol{X}_i$ is a vector of background and psychological characteristics, including IQ and personality traits, and G_{ij} is the guessing indicator defined in Equation (A.2).

$F(\theta^D, \theta^G)$ denotes the joint distribution of the random effects, assumed normal and permitted to be correlated. Such correlation explicitly accounts for the possibility that individuals with lower latent deliberation noise may also differ systematically in their propensity to guess. The variables θ^D and θ^G are integrated out in forming the probability. We parameterize the idiosyncratic and persistent variances $\sigma^2_{u^D}$ and $\sigma^2_{u^G}$ as functions of individual background, IQ, and personality traits, while the elements of $\boldsymbol{U}^D$ and $\boldsymbol{U}^G$ are assumed to be mutually independent. These heteroskedastic structures capture how unobserved behavioral noise varies across individuals and tasks. We assume that agents are independent across i. It is possible to model group-level or school-level effects as additional random clusters, but we do not do so in this paper.

$\boldsymbol{\Omega}$ represents the vector of the following parameters: (i) r in the Expo-Power utility function; (ii) β in the Expo-Power utility function (constrained to zero in models with CRRA utility); (iii) the variance of the transitory deliberation shock $\sigma^2_{u^D}$; (iv) the variance ratio of the persistent to transitory components of the deliberation noise, $\sigma^2_{\theta^D}/\sigma^2_{u^D}$; (v) the variance ratio of the persistent to transitory components of the guessing noise, $\sigma^2_{\theta^G}/\sigma^2_{u^G}$, with $\sigma^2_{u^G}$ normalized to one; (vi) the correlation between the *persistent* components of the deliberation and guessing noises, $\rho(\theta^D, \theta^G)$; and (vii) the guessing index η as in Equation (A.2).[25]

A.3 Identification

Apesteguia and Ballester (2018) show that CRRA models without heteroskedasticity in error variances, guessing behavior, and correlation across choices may generate non-monotonic choice probabilities for risky choices. In that special case, multiple values of a CRRA parameter may be consistent with the same probability of choice for a particular item on a choice list. This possibility leads them and Jagelka (2024) to adopt a random parameter choice model, which builds on the random-utility framework of Thurstone (1927) and Domencich and McFadden (1975), to avoid this possibility. Jagelka's particular specification assumes parameters of preference functions associated with observables that fluctuate

[25] To facilitate smoother estimation, we re-parameterized (iii), (iv), and (v) using log transformations, and (vi) using a hyperbolic tangent transformation. We report the inverse-transformed estimates using the delta method.

randomly across choices, item by item, so that the CRRA risk aversion parameter fluctuates choice by choice. We avoid making this strong assumption in our specification and still, at least locally, identify the model.

We rely on four sources of variation to identify the model: (i) variations in rewards and probabilities across items and choice lists, (ii) inconsistent choice patterns, (iii) cross-individual variation in background and psychological traits, and (iv) functional form restrictions. We establish identification sequentially, following the limit logic of Heckman and Vytlacil (2007). With sufficient variation in arguments of $\boldsymbol{X}$, we can eliminate components of the models. For example, sufficient variation in some traits (e.g., Conscientiousness) can shut down the guessing probability, enabling identification of the rest of the full model. This follows the analyses of Heckman and Vytlacil (2007) and Carneiro et al. (2003). Preference parameters are identified from switching patterns across tasks. Curvature r is pinned down by variation in rewards across choices. In the limit, as the winning probability of a choice approaches one or zero, the deliberation choice becomes degenerate, facilitating identification of other components of the model. In particular, the identification of the β parameter in the Expo-Power model, which represents the wealth-dependence of relative risk aversion, is enabled by the broad range of stake sizes (from 1 to 200 tokens) and by variations in w that we study. Transitory deliberation noise σ_{u^D} is identified from within-person, across-choice variability that cannot be rationalized by stable preferences. ICC^D and ICC^G are identified from correlations in choices across items. The correlation $\rho = \text{Corr}(\theta^D, \theta^G)$ is identified from variation in choices and guesses. Guessing probabilities are identified from patterns of choices inconsistent with the Thurstone model.

As a numerical check of identification, Table A.1 reports the condition numbers of the Hessian matrix of the log-likelihood for each specification, separately for boys and girls, and for both the Expo-Power and CRRA preference models. Across all specifications, the condition numbers range from roughly 10^3 to 10^5, with the corresponding smallest eigenvalues all bounded away from zero (above 10^{-3})—a computational counterpart to local identification, since a non-singular Hessian is what local identification requires (Rothenberg 1971). Larger condition numbers, most visible for girls under Expo-Power without ICC, signal that some directions of the parameter space are weakly determined. The full specification with ICC and guessing is better conditioned for both genders, indicating that the curvature of the likelihood improves once the error structure is correctly specified.

The interpretation of w as a reference wealth level is well defined for adults. For children, it is less so. We perform a robustness analysis that varies w, and our main results are not much affected. In a companion paper (Chávez et al. 2026), we model w explicitly as a function of family background and traits.

Table A.1: Condition Number of Hessian of the Log-Likelihood Across Specifications

	Condition Number of Hessian of $\ln L$ (1)	Smallest Eigenvalue in Hessian of $\ln L$ (2)
Boys, Expo-Power		
With ICC and Guessing	14,816	0.0188
No ICC, with Guessing	35,205	0.0222
With ICC, No Guessing	9,071	0.0051
No ICC, No Guessing	18,110	0.0080
Boys, CRRA		
With ICC and Guessing	7,610	0.0286
No ICC, with Guessing	8,833	0.0218
With ICC, No Guessing	2,476	0.0045
No ICC, No Guessing	1,080	0.0019
Girls, Expo-Power		
With ICC and Guessing	32,735	0.0214
No ICC, with Guessing	240,635	0.0430
With ICC, No Guessing	40,479	0.0091
No ICC, No Guessing	59,334	0.0095
Girls, CRRA		
With ICC and Guessing	5,239	0.0234
No ICC, with Guessing	38,698	0.1048
With ICC, No Guessing	2,542	0.0053
No ICC, No Guessing	1,089	0.0023

Notes: Column (1) reports the ratio of the largest to smallest eigenvalue of the estimated Hessian. Column (2) reports the smallest eigenvalue. All models include IQ and personality traits in each parameter.

B Risk Preference Elicitation Protocol

Risk preferences were elicited using incentivized multiple price list (MPL) tasks administered on tablet computers during the 2019 survey wave. Each fifth-grade student completed six MPL sheets, with choices made through an interactive game interface designed for children aged 10–11. All tasks involved real payoffs, with tokens earned converted to physical prizes at the end of the session.

B.1 Experimental Design and Administration

B.1.1 General Protocol

Before the games began, students were told that after all games were completed, one decision problem would be randomly selected, and that the selected problem would determine how the student's gift was calculated.

- For time-preference games, the gift and its delivery time (e.g., today, two weeks later, one month later) are determined by the student's choice in the selected problem.
- For the other preference games (risk, ambiguity, altruism, trust, reciprocity), the gift is determined by the student's outcome in the selected game according to that game's own rules, and is distributed on the same day.

In all cases, more coins mean better gifts. Assistants showed students that the highest-tier prize consisted of items such as file folders, the second-tier prize of items such as sticky notes, and the third-tier prize of items such as pencils.

B.1.2 Payment Mechanism

Incentives were delivered through a single randomly selected decision. After the student completed the full battery of preference games, one decision problem was drawn at random from across all games, and the student's gift was determined by that problem alone. Because any decision could be the one that counts, this random-selection mechanism makes each choice incentive compatible. In particular, a student's risk-preference gift is realized only when the selected problem happens to be one of the risk MPL items.

B.2 Task Structures: Multiple Price Lists

Students completed six MPL sheets with varying probability structures and payoff ranges. Each sheet presented binary choices between a lower-variance option (Option A) and a higher-variance option (Option B). At every choice item other than the two dominance checks (MPL 2 item 10 and MPL 6 item 10), Option B carries strictly greater payoff variance than Option A. In the certainty-equivalent sheets (MPL 1 and MPL 5) the lower-variance option becomes relatively more attractive as one moves down the list, whereas in the probability-variation sheets (MPL 2–4 and MPL 6) the higher-variance option becomes relatively more attractive as the probability of its high payoff rises.

B.2.1 MPL 1: Standard Certainty Equivalent

This sheet presented choices between a certain amount and a 50–50 lottery with payoffs of 20 or 200 tokens. The certain amount increased from 30 to 130 tokens across items.

Table B.1: MPL 1: Certain Amount vs. 50–50 Lottery

Item	Option A	Option B
1	30 for sure	20 (50%), 200 (50%)
2	40 for sure	20 (50%), 200 (50%)
3	50 for sure	20 (50%), 200 (50%)
4	60 for sure	20 (50%), 200 (50%)
5	70 for sure	20 (50%), 200 (50%)
6	80 for sure	20 (50%), 200 (50%)
7	90 for sure	20 (50%), 200 (50%)
8	100 for sure	20 (50%), 200 (50%)
9	110 for sure	20 (50%), 200 (50%)
10	120 for sure	20 (50%), 200 (50%)
11	130 for sure	20 (50%), 200 (50%)

Notes: This sheet elicits the certainty equivalent of a simple 50–50 lottery, a baseline measure of risk attitudes. The expected value of Option B is 110 tokens. A risk-neutral individual is indifferent at item 9 (certain payoff = 110 = EV of Option B) and switches to Option A from item 10 (certain payoff = 120); individuals who dislike variance switch to the certain amount earlier.

B.2.2 MPL 2: Probability Variation (High Stakes)

Both options are lotteries, with the probability of the high payoff increasing from 10% to 100%. Option A offers 50 or 40 tokens; Option B offers 100 or 10 tokens.

Table B.2: MPL 2: Competing Lotteries with Probability Variation (High Stakes)

Item	Option A	Option B
1	50 (10%), 40 (90%)	100 (10%), 10 (90%)
2	50 (20%), 40 (80%)	100 (20%), 10 (80%)
3	50 (30%), 40 (70%)	100 (30%), 10 (70%)
4	50 (40%), 40 (60%)	100 (40%), 10 (60%)
5	50 (50%), 40 (50%)	100 (50%), 10 (50%)
6	50 (60%), 40 (40%)	100 (60%), 10 (40%)
7	50 (70%), 40 (30%)	100 (70%), 10 (30%)
8	50 (80%), 40 (20%)	100 (80%), 10 (20%)
9	50 (90%), 40 (10%)	100 (90%), 10 (10%)
10	50 for sure	100 for sure

Notes: This sheet identifies the probability threshold at which subjects move from the lower-variance lottery (Option A) to the higher-variance lottery (Option B). The two lotteries have equal expected value at a probability of 37.5% (between items 3 and 4); by item 5, Option B's expected value (55 tokens) exceeds Option A's (45 tokens). A risk-neutral individual switches to Option B by item 4, while more variance-averse individuals switch later. Item 10 is a dominance check (100 for sure dominates 50 for sure) and is not a variance comparison.

B.2.3 MPL 3: Probability Variation (Low Stakes)

This sheet has the same probability structure as MPL 2 but with payoffs scaled down by a factor of five. Option A offers 10 or 8 tokens; Option B offers 20 or 2 tokens.

Table B.3: MPL 3: Competing Lotteries with Probability Variation (Low Stakes)

Item	Option A	Option B
1	10 (10%), 8 (90%)	20 (10%), 2 (90%)
2	10 (20%), 8 (80%)	20 (20%), 2 (80%)
3	10 (30%), 8 (70%)	20 (30%), 2 (70%)
4	10 (40%), 8 (60%)	20 (40%), 2 (60%)
5	10 (50%), 8 (50%)	20 (50%), 2 (50%)
6	10 (60%), 8 (40%)	20 (60%), 2 (40%)
7	10 (70%), 8 (30%)	20 (70%), 2 (30%)
8	10 (80%), 8 (20%)	20 (80%), 2 (20%)
9	10 (90%), 8 (10%)	20 (90%), 2 (10%)

Notes: This sheet tests for stake-size effects by replicating the probability structure of MPL 2 with payoffs scaled down by a factor of five. As in MPL 2, the two lotteries have equal expected value at a probability of 37.5% (between items 3 and 4); at item 5, Option A's expected value is 9 tokens and Option B's is 11 tokens. Comparing switching behavior here with MPL 2 identifies whether risk attitudes depend on the size of the stakes.

B.2.4 MPL 4: Probability Variation (Medium Stakes)

This probability-varying sheet uses intermediate payoff levels. Option A offers 40 or 30 tokens; Option B offers 130 or 10 tokens.

Table B.4: MPL 4: Competing Lotteries with Probability Variation (Medium Stakes)

Item	Option A	Option B
1	40 (10%), 30 (90%)	130 (10%), 10 (90%)
2	40 (20%), 30 (80%)	130 (20%), 10 (80%)
3	40 (30%), 30 (70%)	130 (30%), 10 (70%)
4	40 (40%), 30 (60%)	130 (40%), 10 (60%)
5	40 (50%), 30 (50%)	130 (50%), 10 (50%)
6	40 (60%), 30 (40%)	130 (60%), 10 (40%)
7	40 (70%), 30 (30%)	130 (70%), 10 (30%)
8	40 (80%), 30 (20%)	130 (80%), 10 (20%)
9	40 (90%), 30 (10%)	130 (90%), 10 (10%)

Notes: This sheet adds variation in stakes and expected values to identify potential non-linearities in risk preferences. The two lotteries have equal expected value at a probability of about 18% (between items 1 and 2); from item 2 onward Option B has the higher expected value (e.g. at item 3, Option A = 33 tokens and Option B = 46 tokens). The wide spread in Option B (130 vs. 10) increases the payoff variance relative to Option A while offering a higher mean over most of the list.

B.2.5 MPL 5: Endowment Framing with Gain–Loss Asymmetry

Students began with an endowment of 20 tokens. Option A keeps the endowment; Option B is a 50–50 gamble that adds 20 tokens (a gain) or subtracts a loss L from the endowment, where L increases across items through $\{7, 10, 11, 12, 14, 16, 19\}$. Equivalently, in final-wealth terms Option A pays 20 for sure and Option B pays 40 or $20 - L$ with equal probability.

Table B.5: MPL 5: Endowment Gamble with Varying Loss Magnitude

Item	Option A	Option B	Final-wealth Option B
1	keep 20	endowment 20; win 20 (50%), lose 7 (50%)	40 (50%), 13 (50%)
2	keep 20	endowment 20; win 20 (50%), lose 10 (50%)	40 (50%), 10 (50%)
3	keep 20	endowment 20; win 20 (50%), lose 11 (50%)	40 (50%), 9 (50%)
4	keep 20	endowment 20; win 20 (50%), lose 12 (50%)	40 (50%), 8 (50%)
5	keep 20	endowment 20; win 20 (50%), lose 14 (50%)	40 (50%), 6 (50%)
6	keep 20	endowment 20; win 20 (50%), lose 16 (50%)	40 (50%), 4 (50%)
7	keep 20	endowment 20; win 20 (50%), lose 19 (50%)	40 (50%), 1 (50%)

Notes: This sheet uses an explicit endowment to test for gain–loss asymmetries and reference-dependent preferences: the downside of the higher-variance option deepens across items while the sure amount (keep 20) and the upside (40) are held fixed. Option B's expected value, $30 - L/2$, exceeds the sure 20 for every L in the table (ranging from 26.5 tokens at item 1 to 20.5 tokens at item 7), so a risk-neutral individual always prefers Option B. The item at which a subject abandons Option B for the certain amount identifies the strength of variance and loss aversion, with more averse individuals abandoning it earlier (smaller losses). The final low outcome at item 6 is $20 - 16 = 4$ tokens.

B.2.6 MPL 6: Probability Variation with Endowment (Extended Range)

The final probability-varying sheet also uses the 20-token endowment. In final-wealth terms, Option A offers 60 or 50 tokens; Option B offers 125 or 10 tokens (built as endowment 20 plus win 40/30 for Option A, and endowment 20 plus win 105 or lose 10 for Option B).

Table B.6: MPL 6: Competing Lotteries with Probability Variation (Extended Range)

Item	Option A	Option B
1	60 (10%), 50 (90%)	125 (10%), 10 (90%)
2	60 (20%), 50 (80%)	125 (20%), 10 (80%)
3	60 (30%), 50 (70%)	125 (30%), 10 (70%)
4	60 (40%), 50 (60%)	125 (40%), 10 (60%)
5	60 (50%), 50 (50%)	125 (50%), 10 (50%)
6	60 (60%), 50 (40%)	125 (60%), 10 (40%)
7	60 (70%), 50 (30%)	125 (70%), 10 (30%)
8	60 (80%), 50 (20%)	125 (80%), 10 (20%)
9	60 (90%), 50 (10%)	125 (90%), 10 (10%)
10	60 for sure	125 for sure

Notes: Final-wealth payoffs are shown (each equals the 20-token endowment plus the sheet's gain/loss). This sheet provides a final consistency check and extends the probability and payoff range. The two lotteries have equal expected value at a probability of about 38% (between items 3 and 4); at item 6, Option A's expected value is 56 tokens and Option B's is 79 tokens. Item 10 is a dominance check (125 for sure dominates 60 for sure) and is not a variance comparison. The higher stakes and wider payoff spread add identifying power for the risk-preference parameters.

B.3 Game Formats and Interface Design

The six MPL sheets were delivered through one of two visual interfaces. Each student was assigned a single interface and completed all six sheets in it; the underlying choices are identical across the two interfaces, so the assignment is a between-subjects framing manipulation rather than a difference in the decision problems.

B.3.1 Happy Ball Pool Interface

In the "Happy Ball Pool" interface, each binary choice was presented as a selection between two boxes (A and B) containing colored balls labeled with token payoffs, with the ball distribution representing the probability structure. For example, a 50–50 lottery between 20 and 200 tokens displayed five balls labeled "20" and five labeled "200." On-screen instructions stated: *"The number on the ball represents the number of tokens that you can get. First, you need to choose a box. Then, you can get one ball from the box you* *chose randomly. The number of balls with different numbers will change. Choose a box now!"* After a student tapped a box, the computer drew one ball at random through an animation, and the student received the corresponding tokens. See Figures B.1 and B.2.

Figure B.1: Happy Ball Pool interface: instruction screen.

Figure B.2: Happy Ball Pool interface: choice screen.

B.3.2 Lucky Tokens Interface

In the "Lucky Tokens" interface, the same choices were framed around a 20-token endowment shown as a piggy bank. Students chose whether to keep the tokens (Option A) or "play a game" (Option B) that could add or remove tokens. On-screen rules stated: *"You now have 20 tokens, and they all belong to you. You can choose to keep your tokens safely, or use them to play a game. If you choose to play with me, you can pick one ball from me. You may earn more tokens, or lose part of your tokens."* The final payoff equals 20 plus the number drawn. Because the two interfaces present identical lotteries, comparing behavior across the (randomly assigned) interfaces tests whether the endowment/gain–loss framing shifts measured risk attitudes relative to the neutral "ball pool" framing. See Figures B.3 and B.4.

Figure B.3: Lucky Tokens interface: endowment framing.

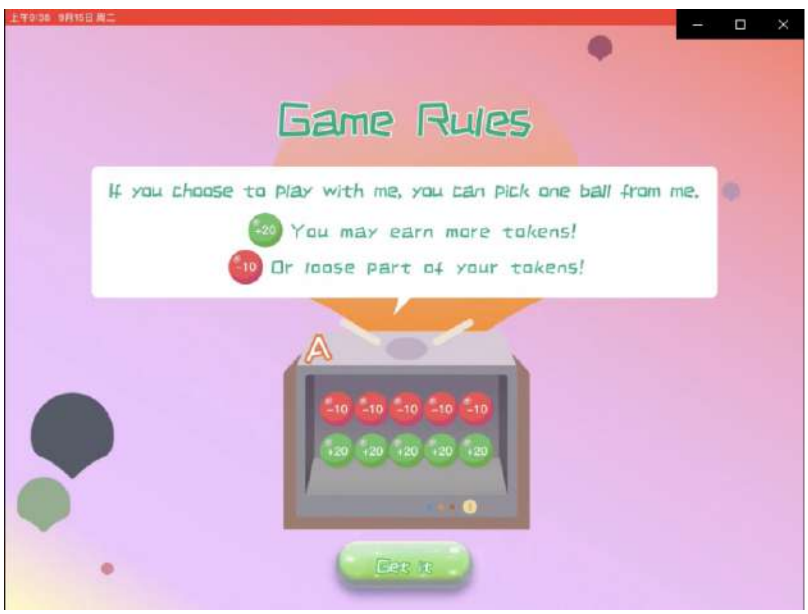


Figure B.4: Lucky Tokens interface: gain–loss display.

B.4 Identifying Risk Preferences from MPL Responses

B.4.1 Switching Points

Under expected utility, a risk-neutral individual switches between the lower- and higher-variance option at the item where the ranking of their expected values reverses. The direction of the switch depends on the sheet: in the certainty-equivalent sheets (MPL 1 and MPL 5) the lower-variance option becomes more attractive down the list, so the switch runs from Option B to Option A; in the probability-variation sheets (MPL 2–4 and MPL 6) the higher-variance option becomes more attractive as the probability of its high payoff rises, so the switch runs from Option A to Option B. In either case, switching toward the lower-variance option sooner indicates greater variance aversion, while switching toward the higher-variance option sooner indicates greater tolerance for payoff variance. Our structural model in Section 2 relaxes this sharp-threshold assumption by explicitly incorporating deliberation noise and guessing behavior, separating true risk preferences from deliberation noise and guessing.

B.4.2 Consistency Checks

Multiple switching points within a single sheet—where a student switches between Option A and Option B more than once—indicate inconsistent choice patterns. Rather than treating these as uninformative or excluding inconsistent respondents, our structural model accounts for them through deliberation noise (σ_{u^D}) and guessing, as described in Section 2. This decomposes observed switching into three components: true preference heterogeneity, persistent individual differences in deliberation noise (captured through the ICC structure), and guessing.

Table E.8 presents reduced-form analysis documenting correlations between inconsistent behavior and observable characteristics. Higher-IQ students and those scoring higher on Conscientiousness exhibit fewer multiple switches. These correlations motivate our structural approach: rather than discarding inconsistent choices, we model them as outcomes of systematic individual differences in deliberation noise and guessing.

B.4.3 Cross-Task Consistency

The six sheets were designed with overlapping regions to test within-subject consistency across contexts. MPL 2, MPL 3, MPL 4, and MPL 6 use probability-variation designs at different stake sizes, enabling tests for stake-size effects; MPL 2 and MPL 3 in particular differ only by a factor-of-five scaling. MPL 1 and MPL 5 both elicit certainty equivalents but under different frames—MPL 1 as simple lotteries over final wealth, MPL 5 with an explicit 20-token endowment and gain/loss structure. Our structural estimation pools information across all six sheets simultaneously, estimating individual-level risk-preference parameters while accounting for sheet-specific features and allowing flexible (Expo-Power) utility that accommodates wealth-dependent risk aversion. Section 5 presents evidence supporting persistent individual traits, with ICC estimates indicating that 38–43% of deliberation-noise variance and 53–58% of guessing variance reflect stable individual traits rather than transitory, sheet-specific shocks.

B.5 Implementation and Data Quality

B.5.1 Training and Standardization

Research assistants received training on the experimental protocol to ensure consistent implementation: memorizing standardized verbal instructions, practicing procedures for answering student questions without influencing choices, technical training on the tablet software, and monitoring for comprehension. Prior to fielding, the software underwent pilot testing with children of similar age not in the main sample; feedback led to refinements in visual design and instruction wording.

B.5.2 Session Length and Fatigue

Students completed all six sheets in a single session. Any rest breaks were built into the interface animations between sheets.

B.5.3 Data Recording

The tablet software automatically recorded the chosen option (A or B), response time in milliseconds, and a timestamp for each item. Data were uploaded to secure servers after each session, and automated checks verified completeness and internal consistency.

B.5.4 Sample Restrictions

Of the 2,204 fifth-grade students in the main analytic sample, all completed all six sheets with no missing data. The overwhelming majority (2,193, 99.5%) exhibited at least one consistent switching pattern across the six sheets, while the remaining 11 (0.5%) showed extreme inconsistency across all sheets. Rather than excluding these "confused" respondents, our structural model accounts for inconsistent choices through its deliberation-noise and guessing components, allowing us to retain the full sample. Excluding the most inconsistent respondents would bias the sample toward less noisy decision-makers and understate the importance of deliberation noise and guessing.

B.6 Methodological Advantages

The design incorporates several features that enhance measurement quality and identification. First, six sheets with varying structures provide rich cross-sectional variation for identifying risk-preference parameters, spanning different probability ranges (from 10% to 100%), stake sizes (from 1 to 200 tokens), and reference frames (simple lotteries versus endowment/gain–loss framing). Second, the child-appropriate interface balances engagement with clarity by representing uncertainty through visual ball distributions. Third, real incentives with random-problem selection ensure incentive compatibility. Fourth, individual administration minimizes peer influence and enables precise measurement of response times. Fifth, the standardized protocol ensures consistency across classrooms and schools. Sixth, the longitudinal design of the Longitudinal Study of Children's Development in Mianzhu enables future analysis of preference stability and development.

These features, combined with the cognitive and personality data described in Section 3 and Appendix C, enable our structural analysis of how individual differences shape both risk preferences and the precision with which they are expressed.

C Measurement Instruments and Psychometric Properties

This appendix documents the measurement instruments used to assess IQ and personality traits in the Longitudinal Study of Children's Development in Mianzhu. We provide details on test administration, psychometric properties, and the factor analytic approach used to construct measurement-error-corrected personality trait estimates.

C.1 IQ: Raven's Progressive Matrices

Students' IQ was measured using the Raven's Standard Progressive Matrices (SPM), a widely used, non-verbal test of fluid intelligence and abstract reasoning ability. The SPM is particularly appropriate for cross-cultural research as it minimizes language and cultural bias while providing a reliable measure of IQ.

The Raven's SPM consists of 60 items arranged in five sets (A through E) of 12 items each. Each item presents an incomplete matrix of geometric patterns with one element missing, and students select which of 6–8 response options correctly completes the pattern. Items increase in difficulty within and across sets, progressing from simple pattern completion in Set A to complex analytic reasoning in Set E.

In the Longitudinal Study of Children's Development in Mianzhu, the complete 60-item battery was administered by paper and pencil to students in Grades 4, 5, and 6 from the 18 primary schools in our sample during waves 2017–2020. The test was administered at the same time as the survey under standardized conditions and, in each class, was proctored by the head teacher together with study interviewers (Feng et al. 2022). Students were given 40 minutes to complete the test.

We estimate the standardized IQ index for Grade 5 in 2019 using an IRT-3PL model and standardize it to have mean 0 and standard deviation 1. The IQ index for other grade-waves is estimated using an IRT-3PL model in which the discrimination, difficulty, and guessing parameters for each test item are fixed at their estimates from the Grade 5 (2019) sample. This approach leverages the fact that the same 60 items are used across grade-waves and accounts for the growth of IQ over time.

C.2 Personality Traits: Big Five Inventory-2

Personality traits were measured using an abbreviated form of the Big Five Inventory-2 (BFI-2), developed by Soto and John (2017). The BFI-2 is a widely validated instrument that assesses the five major dimensions of personality: Conscientiousness, Agreeableness, Extraversion, Openness, and Emotional Stability.

The full BFI-2 consists of 60 items, 12 per dimension, designed to capture the breadth of each personality domain. To limit respondent burden, the Longitudinal Study of Children's Development in Mianzhu administers an abbreviated instrument that retains four items per dimension (Feng et al. 2022; Feng et al. 2026). Our main analysis uses the teacher report, which comprises 20 items (four for each of the five domains); a longer student self-report (62 items) was also collected but is not used in our structural estimation, for the reliability reasons discussed below.

Items are short phrases describing behaviors, thoughts, or feelings (e.g., "Is helpful and unselfish with others," "Is outgoing, sociable"). Teachers rated how well each statement describes each student using a 5-point Likert scale ranging from 1 (Disagree strongly) to 5 (Agree strongly).

We use a validated Chinese translation of the BFI-2, adapted to be appropriate for primary-school children (Feng et al. 2022). The translation process followed standard back-translation procedures, and the Chinese version demonstrates comparable psychometric properties to the original English version. In our main analysis, we focus on Grade 5 (2019), when personality traits were assessed through teacher reports, which demonstrate superior reliability and predictive power compared to student self-assessments (Feng et al. 2022). Recent research using the Longitudinal Study of Children's Development in Mianzhu confirms that these teacher-reported personality measures respond to environmental inputs during childhood, including the classroom peer composition to which children are randomly assigned (Feng et al. 2026).

C.3 Factor Score Estimation: Multiple-Indicator Measurement System

Our personality trait measures are constructed using factor analysis with explicit measurement error correction. Rather than using simple item averages, we account for heterogeneity in item reliability and construct optimal weighted factor scores.

For each Big Five dimension, we specify a four-indicator measurement system in which the observed teacher-rated item responses M_j ($j = 1, 2, 3, 4$) load on a common latent factor θ representing the true underlying trait. The measurement equations are:

$$M_1 = \theta + \eta_1 \tag{C.1}$$

$$M_2 = \nu_2 + \phi_2\theta + \eta_2 \tag{C.2}$$

$$M_3 = \nu_3 + \phi_3\theta + \eta_3 \tag{C.3}$$

$$M_4 = \nu_4 + \phi_4\theta + \eta_4 \tag{C.4}$$

where the first equation is normalized with $\nu_1 = 0$ and $\phi_1 = 1$ for identification, and η_j are item-specific measurement errors satisfying $E(\eta_j) = 0$ for all j, $\eta_j \perp\!\!\!\perp \theta$ (measurement errors independent of the true trait), and $\mathrm{Cov}(\eta_k, \eta_{k'}) = 0$ for $k \neq k'$ (uncorrelated measurement errors across items). Prior to estimation, reverse-keyed items (e.g., "Often feels sad," "Leaves a mess or does not clean up," "Introverted, does not love socializing") are recoded so that all four indicators of a dimension are positively keyed.

The system (C.1)–(C.4) separates the observed item responses into two components: the common factor θ reflecting true trait variation, and the item-specific errors η_j capturing measurement noise. The factor loadings ϕ_j determine how strongly each item relates to the underlying trait, while the intercepts ν_j allow for item-specific response tendencies.

C.3.1 Identification and Estimation of Model Parameters

With four indicators, the sample covariance structure supplies ten distinct second moments (four variances and six covariances) to identify eight parameters (three loadings, four error variances, and the factor variance), so the one-factor model is over-identified with two degrees of freedom. The loadings and error variances are identified from covariances among the items. For intuition, any triple of indicators (M_1, M_j, M_k) identifies the loading on M_j as a ratio of covariances,

$$\phi_j = \frac{\mathrm{Cov}(M_j, M_k)}{\mathrm{Cov}(M_1, M_k)}, \qquad j, k \in \{2, 3, 4\},\ j \neq k, \tag{C.5}$$

the factor mean and variance as

$$E(\theta) = E(M_1) = \bar{M}_1, \tag{C.6}$$

$$\mathrm{Var}(\theta) = \frac{\mathrm{Cov}(M_1, M_j)}{\phi_j}, \tag{C.7}$$

and each error variance as the residual after netting out the common factor,

$$\mathrm{Var}(\eta_k) = \mathrm{Var}(M_k) - \phi_k^2 \,\mathrm{Var}(\theta). \tag{C.8}$$

Because each trait has four indicators, one more than the three required for identification, each of these objects is over-determined: distinct triples in (C.5) and distinct choices of j in (C.7) yield algebraically equivalent expressions at the population level but distinct sample estimators. Rather than select one, we estimate all parameters jointly by maximum likelihood, which optimally combines the over-identifying moments and delivers the confirmatory-factor-analysis estimates reported in Table C.1.

C.3.2 Bartlett GLS Estimator of Factor Scores

Given the identified parameters, we construct individual-level trait estimates using the Bartlett (1937) GLS estimator, which provides the minimum mean squared error predictor of the latent factor. For person i, the Bartlett factor score is:

$$\hat{\theta}_i^B = \left[\boldsymbol{\phi}'\boldsymbol{\Omega}^{-1}\boldsymbol{\phi}\right]^{-1}\left[\boldsymbol{\phi}'\boldsymbol{\Omega}^{-1}\left(\boldsymbol{M}_i - \boldsymbol{\nu}\right)\right] \tag{C.9}$$

where $\boldsymbol{\phi} = (1, \phi_2, \phi_3, \phi_4)'$ is the vector of factor loadings, $\boldsymbol{\nu} = (0, \nu_2, \nu_3, \nu_4)'$ is the vector of item intercepts, $\boldsymbol{M}_i = (M_{i1}, M_{i2}, M_{i3}, M_{i4})'$ is person i's vector of observed item responses, and $\boldsymbol{\Omega}$ is the diagonal matrix of measurement error variances:

$$\boldsymbol{\Omega} = \begin{bmatrix} \mathrm{Var}(\eta_1) & 0 & 0 & 0 \\ 0 & \mathrm{Var}(\eta_2) & 0 & 0 \\ 0 & 0 & \mathrm{Var}(\eta_3) & 0 \\ 0 & 0 & 0 & \mathrm{Var}(\eta_4) \end{bmatrix} \tag{C.10}$$

The Bartlett estimator in equation (C.9) optimally weights items according to their reliability. Items with lower uniqueness (higher signal-to-noise ratio) receive greater weight in the factor score, as they provide more precise information about the latent trait. This weighting scheme is optimal in the sense of minimizing the mean squared error of the factor score as a predictor of the true latent trait θ, and it coincides with the reliability-weighting logic used to construct the Bartlett scores in the prior Mianzhu studies (Feng et al. 2022; Feng et al. 2026).

By explicitly modeling measurement error through the four-equation system (C.1)–(C.4) and extracting purged factor scores via equation (C.9), we obtain personality trait measures that are corrected for item-specific measurement noise. These measurement-error-corrected factor scores are subsequently used as covariates in our structural model of risk preferences and deliberation noise (Section 2 of the main text).

C.4 Measurement Quality and Reliability

Table C.1 presents the factor uniqueness (measurement error variance) for each of the 20 Big Five items used in our four-equation systems, based on teacher assessments in Grade 5 (2019). Factor uniqueness represents the proportion of item variance not explained by the latent factor—that is, $\mathrm{Var}(\eta_j)/\mathrm{Var}(M_j)$. Lower values indicate more reliable items with stronger factor loadings. These measurement quality estimates are consistent with Feng et al. (2022), who demonstrate that teacher reports exhibit higher internal consistency than child or guardian reports in the Longitudinal Study of Children's Development in Mianzhu.

Table C.1: Factor Uniqueness (Measurement Errors) of Big Five Items: Teacher Assessments of Grade 5 Students' Personality Traits in 2019

Description	Uniqueness	Description	Uniqueness
Conscientiousness (4)		*Openness (4)*	
Is dependable	0.569	Is curious about many different things	0.647
Leaves a mess or does not clean up	0.600	Is inventive and finds clever ways to do things	0.198
Keeps things neat and tidy	0.269	Is fond of arts	0.748
Is efficient or gets things done	0.617	Is original or comes up with new ideas	0.205
Agreeableness (4)		*Emotional Stability (4)*	
Is compassionate or has a soft heart	0.351	Is relaxed or handles stress well	0.643
Has a forgiving nature	0.205	Is emotionally stable or not easily upset	0.497
Is helpful and unselfish with others	0.244	Often feels sad	0.872
Is polite or courteous to others	0.496	Keeps their emotions under control	0.368
Extraversion (4)			
Has a confident, strong, or bold personality	0.362		
Is outgoing or sociable	0.277		
Introverted, does not love socializing	0.631		
Is dominant or acts as a leader	0.648		

Notes: Factor uniqueness represents the proportion of item variance not explained by the latent factor (i.e., measurement error). Lower values indicate more reliable items. Estimates based on Grade 5 teacher assessments (N = 2,244) from confirmatory factor analysis using the four-equation system in equations (C.1)–(C.4).

Several patterns emerge from the measurement error analysis. First, reliability varies substantially across items. The most reliable item is "Is inventive and finds clever ways to do things" (Openness, uniqueness = 0.198), indicating that approximately 80% of item variance reflects true trait variation. In contrast, "Often feels sad" (Emotional Stability) shows the highest measurement error (uniqueness = 0.872), suggesting that only 13% of variance reflects the underlying trait. This heterogeneity in reliability motivates our use of the Bartlett estimator in equation (C.9), which optimally down-weights unreliable items.

Second, average measurement quality is acceptable. The mean uniqueness across all 20 items is 0.47, indicating that approximately 53% of item variance reflects true personality variation. Within dimensions, Agreeableness items show the highest average reliability (mean uniqueness = 0.32), while Emotional Stability items show the lowest (mean uniqueness = 0.60).

Third, observable behaviors show higher reliability than internal states. Items capturing externally visible characteristics—such as inventiveness, sociability, and task completion—show lower uniqueness (higher reliability) than items related to internal emotional states. For example, "Is outgoing or sociable" (uniqueness = 0.277) is measured more reliably than "Often feels sad" (uniqueness = 0.872). This pattern reflects the inherent difficulty of assessing internal experiences through teacher observation and provides additional justification for our use of teacher reports, as teachers can more accurately rate observable behaviors.

These measurement error estimates directly inform the weighting scheme in our Bartlett estimator (equation C.9). Items with lower uniqueness receive greater weight in constructing factor scores, yielding more precise trait estimates than equal-weighted averages would provide.

C.5 Teacher vs. Student Reports: Validation Analysis

Research using the Longitudinal Study of Children's Development in Mianzhu confirms that teacher reports have superior psychometric properties and are more predictive of cognitive outcomes and school behavior than child or guardian reports (Feng et al. 2022). Teacher reports also better capture meaningful variation in personality traits that respond to environmental factors such as peer influences (Feng et al. 2022).

To validate our use of teacher assessments, we collected both teacher and student self-reports of personality in a subset of 612 sixth-grade students across six schools. Table C.2 presents Spearman rank correlations between the two assessment modes for selected items from each personality dimension.

Table C.2: Spearman Rank Correlations Between Teacher and Student Reports

Trait	Item	Rank Correlation
Conscientiousness	Is dependable, steady	0.17
	Keeps things neat and tidy	0.15
	Is efficient, gets things done	0.19
	Leaves a mess, doesn't clean	0.13
Agreeableness	Is compassionate, has a soft heart	0.09
	Has a forgiving nature	0.06
	Is helpful and unselfish with others	0.10
	Is polite, courteous to others	0.08
Extraversion	Is outgoing, sociable	0.21
	Has an assertive personality	0.19
	Is sometimes shy, introverted	0.19
	Is dominant, acts as a leader	0.26
Openness	Is curious about many different things	0.09
	Is inventive, finds clever ways to do things	0.12
	Values art and beauty	0.22
	Is original, comes up with new ideas	0.14
Emotional Stability	Is relaxed, handles stress well	0.05
	Is emotionally stable, not easily upset	−0.02
	Often feels sad	0.08
	Keeps their emotions under control	0.02

Notes: Spearman rank correlations between teacher and student reports for Grade 6 students in 2019. $N = 612$ students from six schools who completed both assessments.

The correlations between teacher and student reports range from −0.02 to 0.26 across the 20 items assessed. Several patterns emerge. First, correlations vary by trait dimension. Extraversion items show the strongest agreement between teachers and students (range: 0.19–0.26), with "Is dominant, acts as a leader" exhibiting the highest correlation ($\rho_s = 0.26$). In contrast, Emotional Stability items show the weakest correlations (range: −0.02 to 0.08), with "Is emotionally stable, not easily upset" showing essentially zero agreement ($\rho_s = -0.02$).

Second, observable behaviors show higher agreement than internal states. Items capturing externally visible characteristics—such as dominance, sociability, and conscientiousness in completing tasks—show moderate correlations (0.15–0.26). Items related to internal emotional states or abstract dispositions show weaker correlations (0.02–0.09), suggesting that teachers have limited ability to assess children's inner feelings or that children and teachers use different reference frames when evaluating these traits.

Third, overall agreement is modest but positive. The average correlation across all 20 items is 0.12, indicating that teacher and student assessments capture overlapping but distinct information. This pattern is consistent with multi-rater personality research showing that different informants provide partially independent perspectives on traits.

Based on this validation analysis, we use teacher assessments as our primary personality measures throughout the main analysis. Teacher reports offer several advantages for studying elementary school children: teachers observe students across multiple contexts and can compare them to age-appropriate peer benchmarks, reducing the impact of limited self-awareness and reading comprehension difficulties that may affect young children's self-reports (Feng et al. 2022).

C.6 Developmental Patterns in Personality and IQ

Figures C.1–C.5 present descriptive evidence on how personality traits and IQ evolve across Grades 4, 5, and 6, and how these patterns vary across measurement methods and demographic groups.

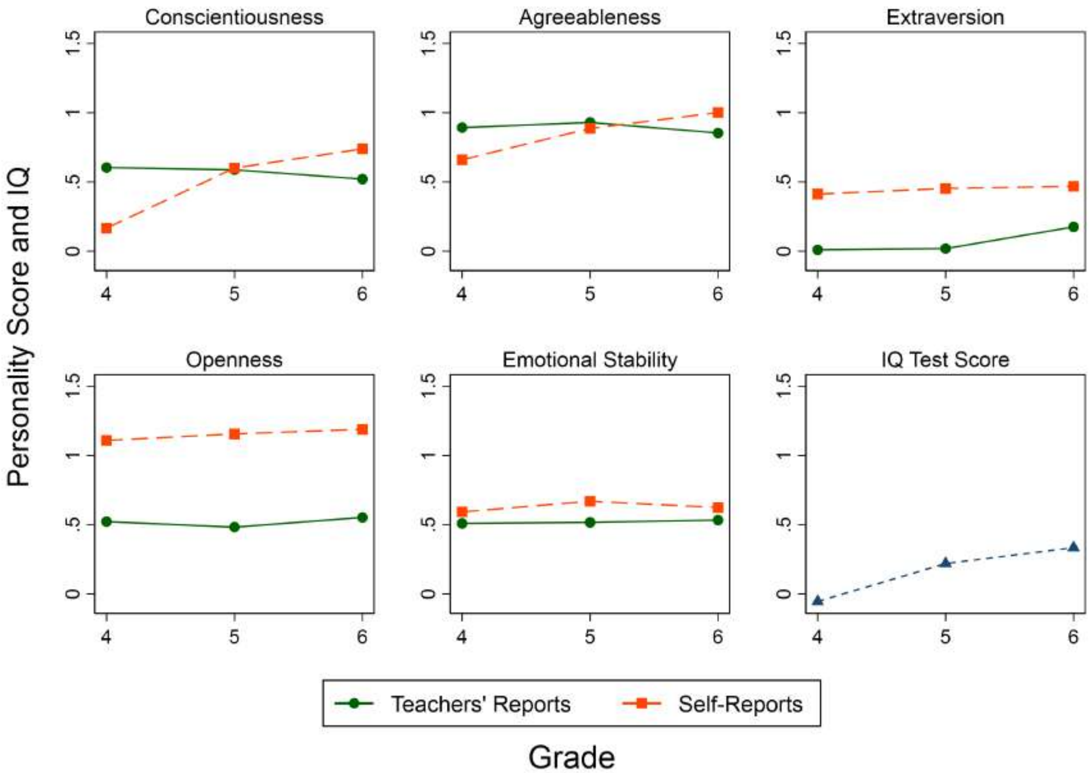


Figure C.1: Personality Traits and IQ Across Grades: Teacher Reports vs. Self-Reports

Notes: Standardized personality scores (Big Five dimensions) and IQ test scores across Grades 4, 5, and 6. Green solid lines show teacher reports, orange dashed lines show student self-reports. Teacher and self-reports diverge substantially for most traits: self-reports exceed teacher reports for Openness and Extraversion at every grade, while for Conscientiousness and Agreeableness teacher reports are higher in the early grades and self-reports rise to cross them around Grade 5. IQ test scores (bottom right panel) increase consistently across grades.

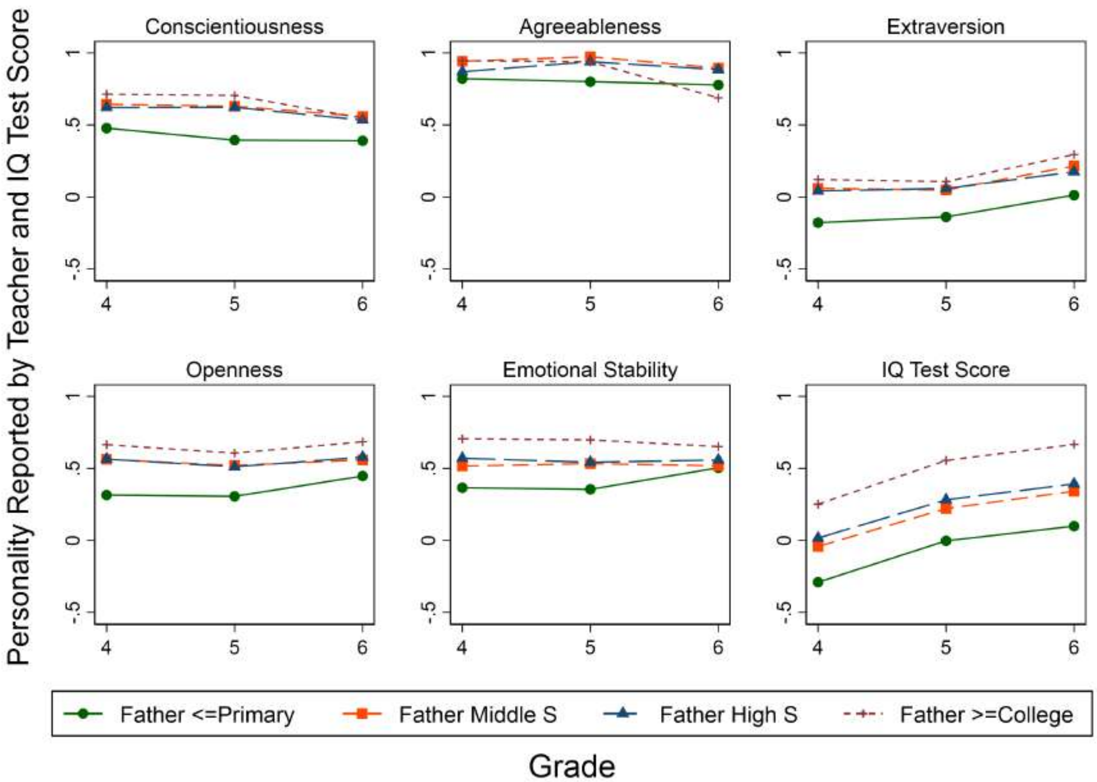


Figure C.2: Personality Traits and IQ by Father's Education Level

Notes: Teacher-reported personality scores and IQ test scores across Grades 4, 5, and 6, stratified by father's education level. Four categories shown: Father ≤Primary (green solid), Father Middle School (orange solid), Father High School (gray solid), and Father ≥College (gray dashed). Clear socioeconomic gradients are visible for IQ test scores, with stronger effects in higher grades. Personality trait differences by parental education are modest.

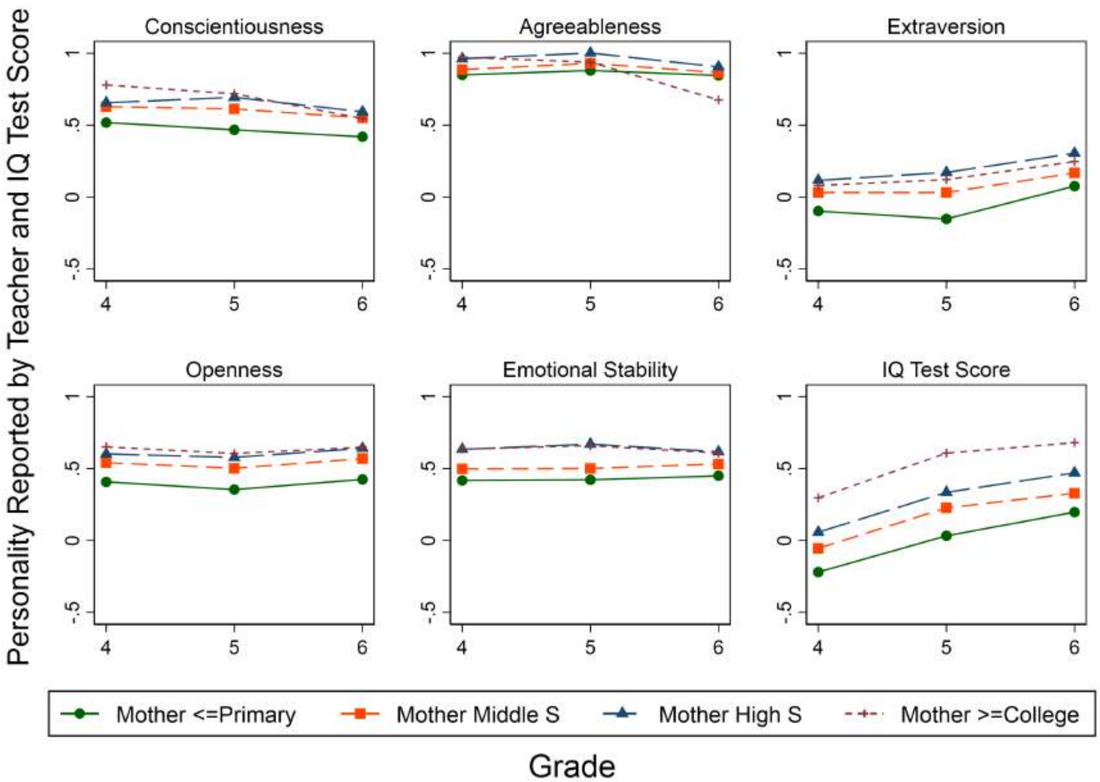


Figure C.3: Personality Traits and IQ by Mother's Education Level

Notes: Teacher-reported personality scores and IQ test scores across Grades 4, 5, and 6, stratified by mother's education level. Four categories shown: Mother ≤Primary (green solid), Mother Middle School (orange solid), Mother High School (gray solid), and Mother ≥College (gray dashed). Socioeconomic gradients in IQ are evident and strengthen with age. Personality traits show less systematic variation by maternal education.

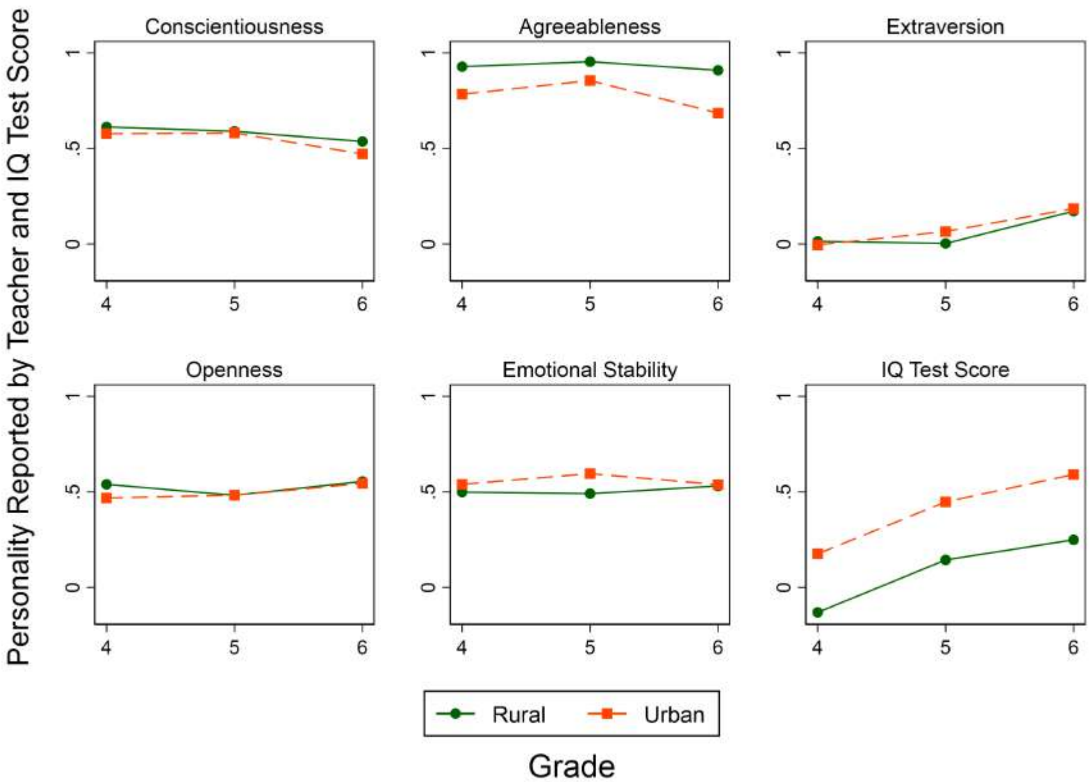


Figure C.4: Personality Traits and IQ: Rural vs. Urban Students

Notes: Teacher-reported personality scores and IQ test scores across Grades 4, 5, and 6, comparing rural (green solid) and urban (orange dashed) students. Urban students show consistently higher IQ scores across all grades, with the gap widening over time. Personality trait differences between rural and urban students are small and inconsistent across traits.

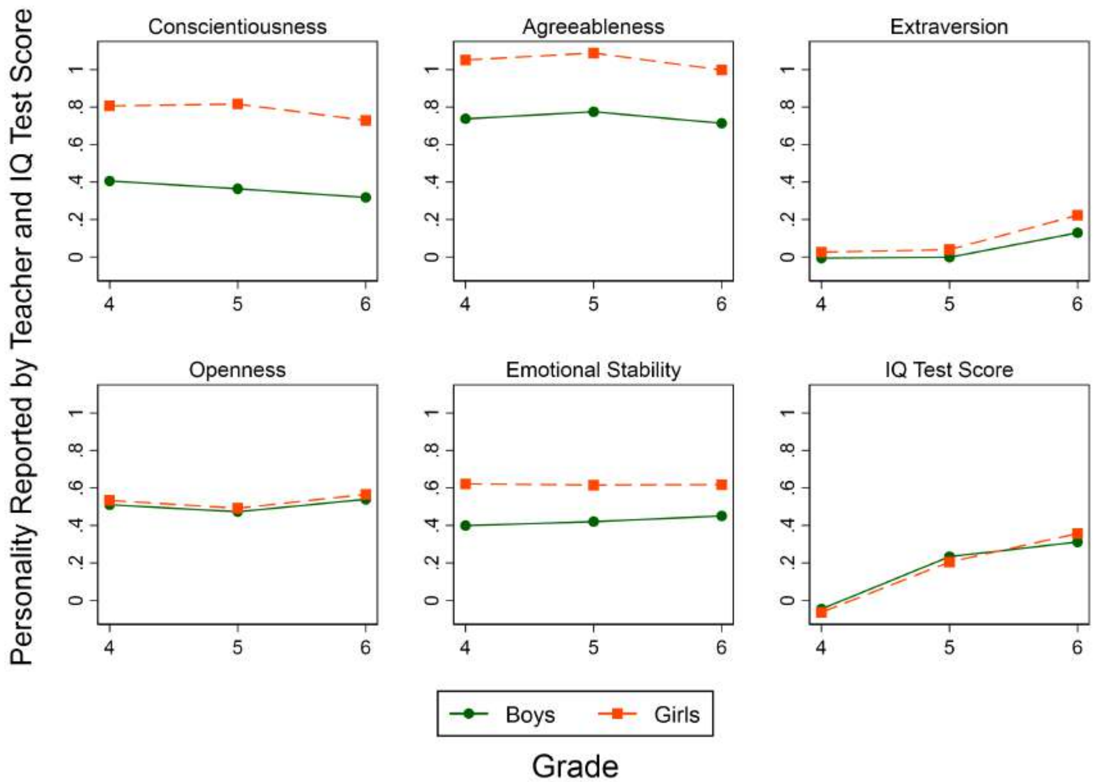


Figure C.5: Gender Differences in Personality Traits and IQ

Notes: Teacher-reported personality scores and IQ test scores across Grades 4, 5, and 6, comparing boys (green solid) and girls (orange dashed). Girls score higher than boys on Conscientiousness, Agreeableness, and Emotional Stability across all grades. Boys and girls show essentially identical IQ at every grade.

Teacher Reports vs. Self-Reports. Figure C.1 compares teacher assessments and student self-reports of personality traits across grades. For Openness and Extraversion, self-reports exceed teacher reports at every grade, so that students rate themselves more favorably on these traits. For Conscientiousness and Agreeableness the pattern reverses over time: teacher reports are high and stable while self-reports begin well below and rise across grades, crossing teacher reports around Grade 5. IQ test scores show steady increases from Grade 4 to Grade 6, consistent with normal cognitive development.

Socioeconomic Gradients. Figures C.2 and C.3 show personality and IQ patterns stratified by parental education. Socioeconomic gradients in IQ are substantial and strengthen with age: students whose parents have college education score approximately 0.5 standard deviations higher than students whose parents have only primary education by Grade 6. In contrast, personality traits show weaker and less consistent associations with parental education, suggesting that family background affects cognitive development more strongly than personality development during this age range.

Rural-Urban Differences. Figure C.4 compares rural and urban students. Urban students exhibit consistently higher IQ scores, by about 0.3 standard deviations, a gap that is stable across Grades 4–6. This disparity likely reflects differences in educational resources between urban and rural settings. Personality trait differences between rural and urban students are modest and do not show consistent patterns across traits or grades.

Gender Differences. Figure C.5 reveals substantial and persistent gender differences in personality traits but a more nuanced pattern for IQ. Girls score higher than boys on Conscientiousness, Agreeableness, and Emotional Stability across all grades, with differences ranging from about 0.2 to 0.4 standard deviations. IQ is essentially identical across genders at all grades (Grades 4–6), consistent with the Grade-5 estimates in Table 1.

These descriptive patterns inform our main analysis in several ways. First, they justify our focus on Grade 5 as a period when personality traits are relatively stable but still in development. Second, they highlight the importance of controlling for socioeconomic background when studying the relationship between traits and preferences. Third, they underscore the need to model gender differences explicitly in our structural estimation framework. Finally, they support our choice of teacher reports over self-reports, as teacher assessments show more plausible variation across demographic groups and avoid the positive self-presentation bias evident in student self-reports.

C.7 Implications for Main Analysis

The measurement properties documented in this appendix inform our structural estimation approach in several important ways. By using the four-equation factor system and Bartlett estimator, we construct personality trait measures explicitly corrected for item-specific measurement error, yielding more accurate trait estimates than simple item averages. Teacher reports demonstrate superior reliability properties across all Big Five dimensions compared to alternative respondent types (Feng et al. 2022), justifying our reliance on teacher assessments. Using purged factor scores eliminates attenuation bias that would arise from measurement error in covariates, ensuring our structural estimates reflect true trait effects rather than contaminated relationships. Evidence from Feng et al. (2022) confirms that teacher reports are substantially more predictive of school outcomes than child or guardian reports, even after accounting for potential rater effects, supporting our decision to use teacher-reported personality measures as covariates in our structural model. Research demonstrates that these teacher-reported personality measures capture meaningful individual differences that respond to environmental inputs and predict important life outcomes (Feng et al. 2022).

In our structural estimation framework, we use the measurement-error-corrected Bartlett factor scores as personality covariates when modeling risk preferences, deliberation noise, and guessing behavior. This ensures that our parameter estimates capture the effects of true underlying personality traits on economic preferences and deliberation noise, free from the attenuation bias that would arise from using contaminated trait measures.

D Teacher-Reported vs. Self-Reported Personality Traits

This appendix examines whether our results depend on the source of personality trait measurement by comparing teacher-reported Big Five (our main specification) to self-reported Big Five. Teachers observe children across multiple contexts over extended periods, potentially providing more objective assessments, while self-reports capture children's internal experiences and self-perceptions. We compare two specifications under both the Expo-Power and CRRA frameworks with full ICC structure and guessing: Model 1 uses teacher-reported traits (our baseline), and Model 2 uses self-reported traits. The comparison reveals both convergence and divergence: the level parameters (σ_{u^D}, ICC^D, ICC^G) remain stable within 5–10%, while the deliberation–guessing correlation ρ is more sensitive to measurement source (teacher -0.15 vs. self -0.19 for boys, and 0.11 vs. 0.16 for girls), and coefficient patterns differ substantially, with self-reports showing weaker and noisier trait-preference relationships. This suggests teacher assessments provide more reliable measurement of traits relevant for economic decision-making, validating our methodological choice.

D.1 Expo-Power Specification

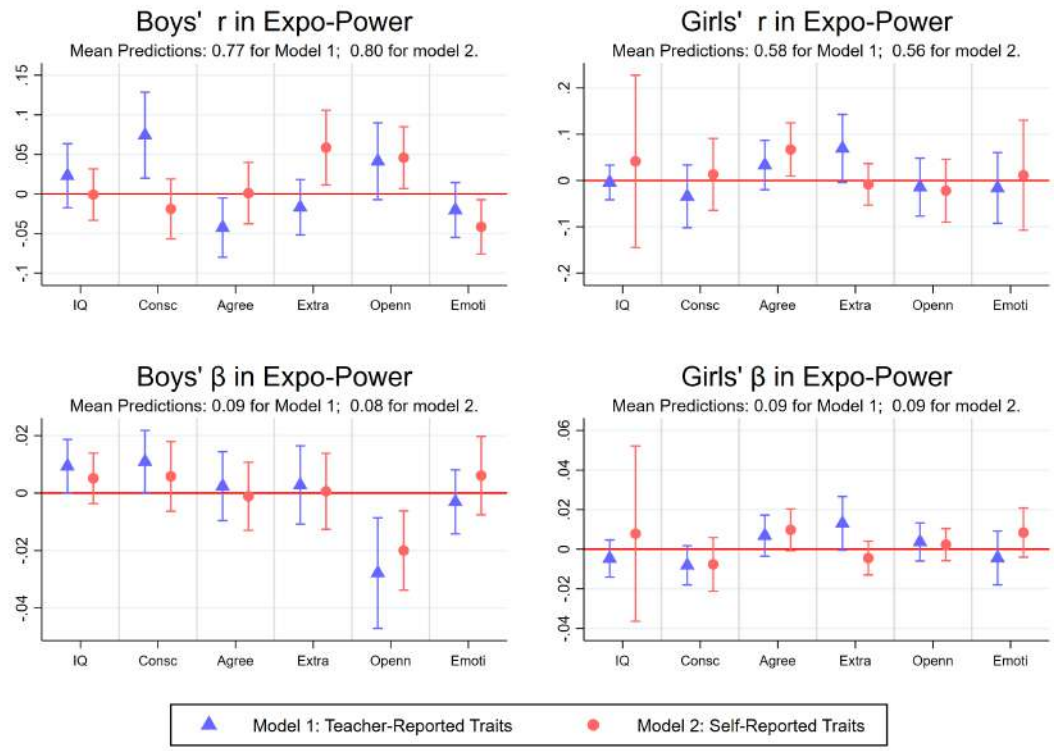


Figure D.1: Teacher vs. Self Reports: Risk Aversion Parameters r and β (Expo-Power)

Notes: We compute predictions for each observation and average them over the sample. Blue triangles: Model 1 (teacher-reported traits, our baseline). Red circles: Model 2 (self-reported traits). Mean predictions appear in each panel's subtitle. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

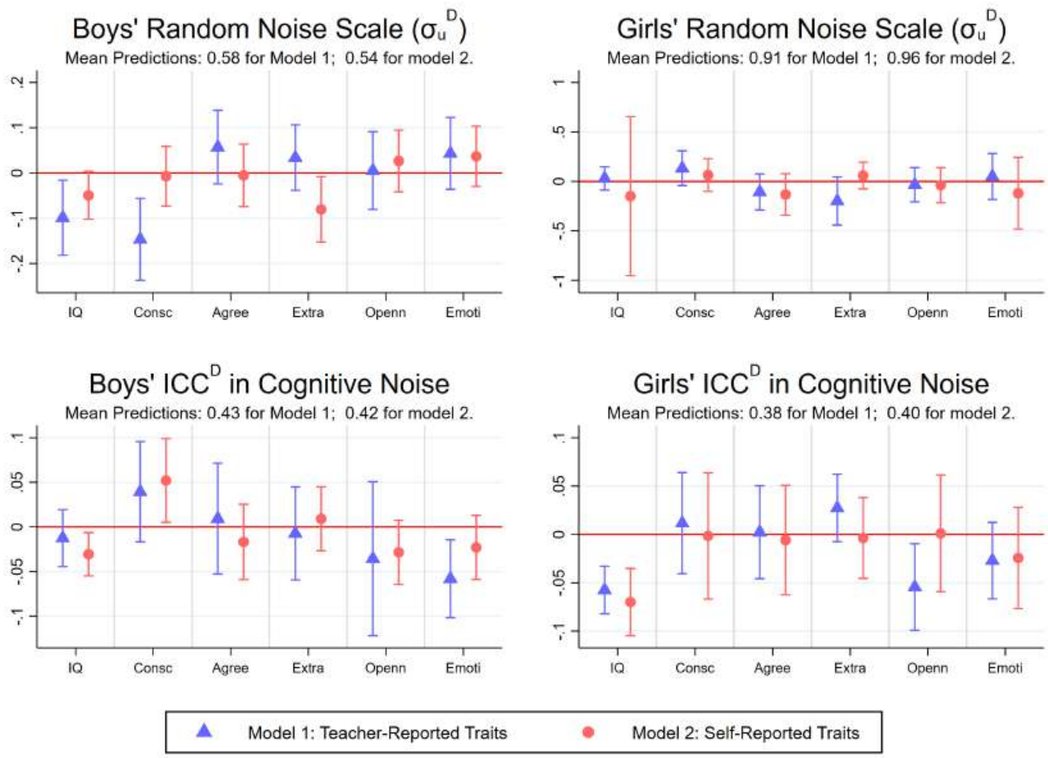


Figure D.2: Teacher vs. Self Reports: Transitory Noise σ_{u^D} and ICC^D (Expo-Power)

Notes: We compute predictions for each observation and average them over the sample. Blue triangles: Model 1 (teacher-reported traits, our baseline). Red circles: Model 2 (self-reported traits). Mean predictions appear in each panel's subtitle. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

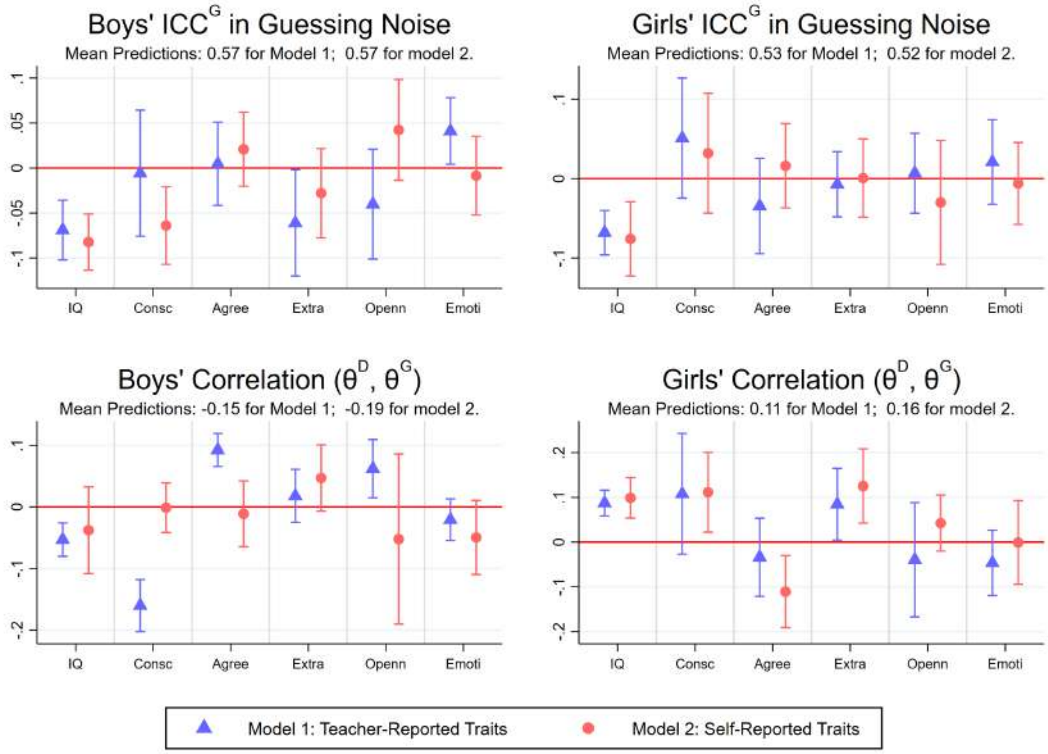


Figure D.3: Teacher vs. Self Reports: Guessing Persistence ICC^G and Deliberation–Guessing Correlation (Expo-Power)

Notes: We compute predictions for each observation and average them over the sample. Blue triangles: Model 1 (teacher-reported traits, our baseline). Red circles: Model 2 (self-reported traits). Mean predictions appear in each panel's subtitle. Error bars show 90% confidence intervals. N = 1,115 boys, N = 1,089 girls.

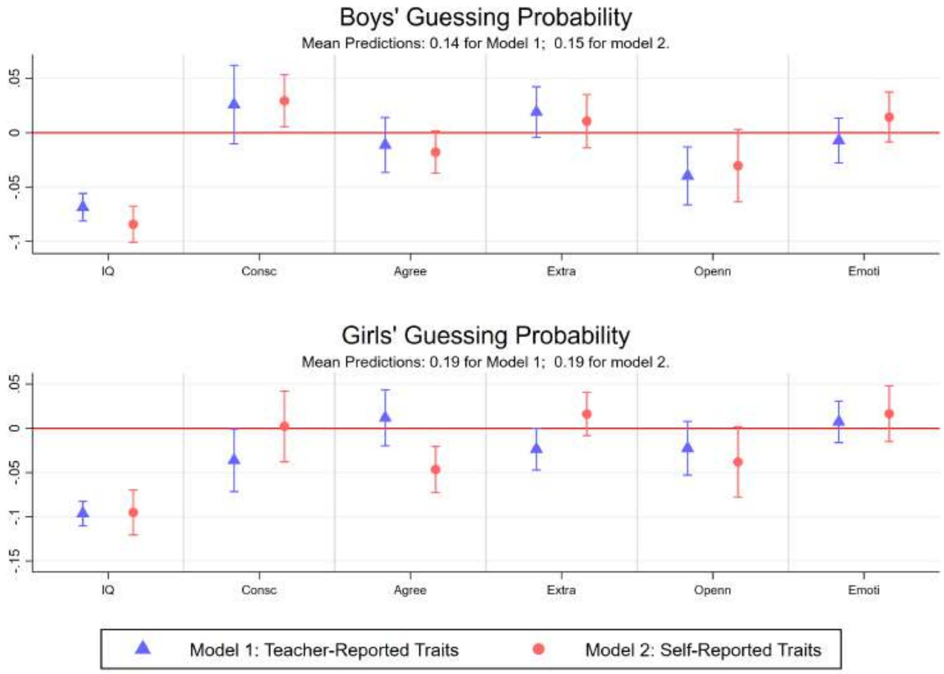


Figure D.4: Teacher vs. Self Reports: Guessing Probability (Expo-Power)

Notes: We compute predictions for each observation and average them over the sample. Blue triangles: Model 1 (teacher-reported traits, our baseline). Red circles: Model 2 (self-reported traits). Mean predictions appear in each panel's subtitle. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

D.2 CRRA Specification

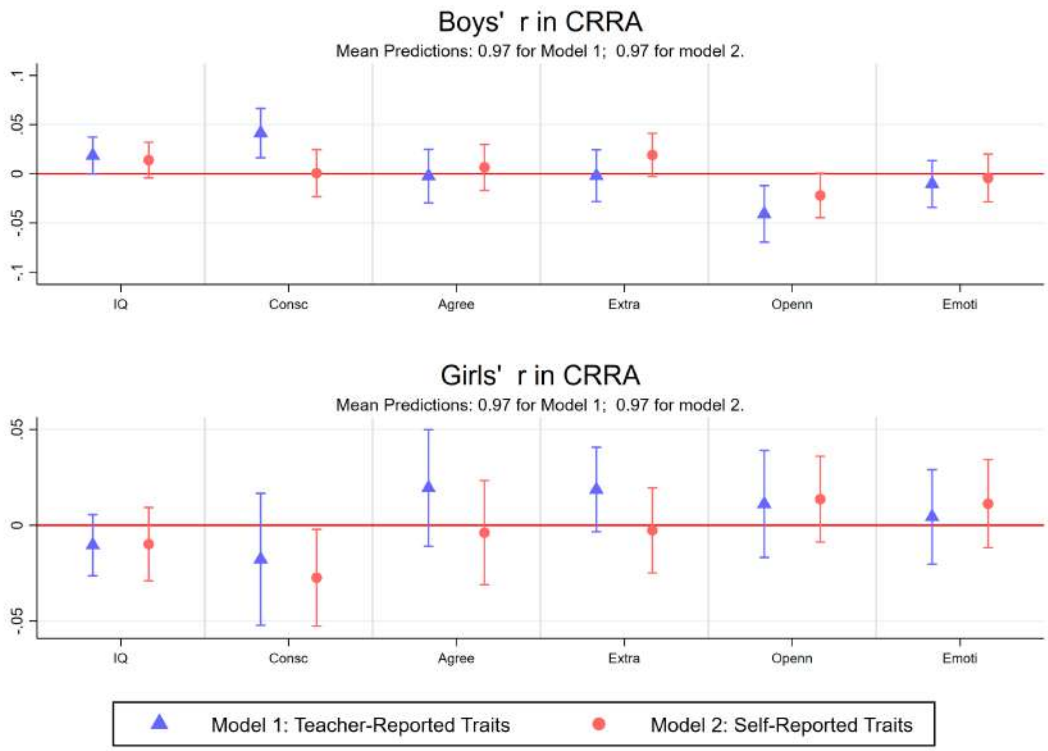


Figure D.5: Teacher vs. Self Reports: Risk Aversion r (CRRA)

Notes: We compute predictions for each observation and average them over the sample. Blue triangles: Model 1 (teacher-reported traits, our baseline). Red circles: Model 2 (self-reported traits). Mean predictions appear in each panel's subtitle. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

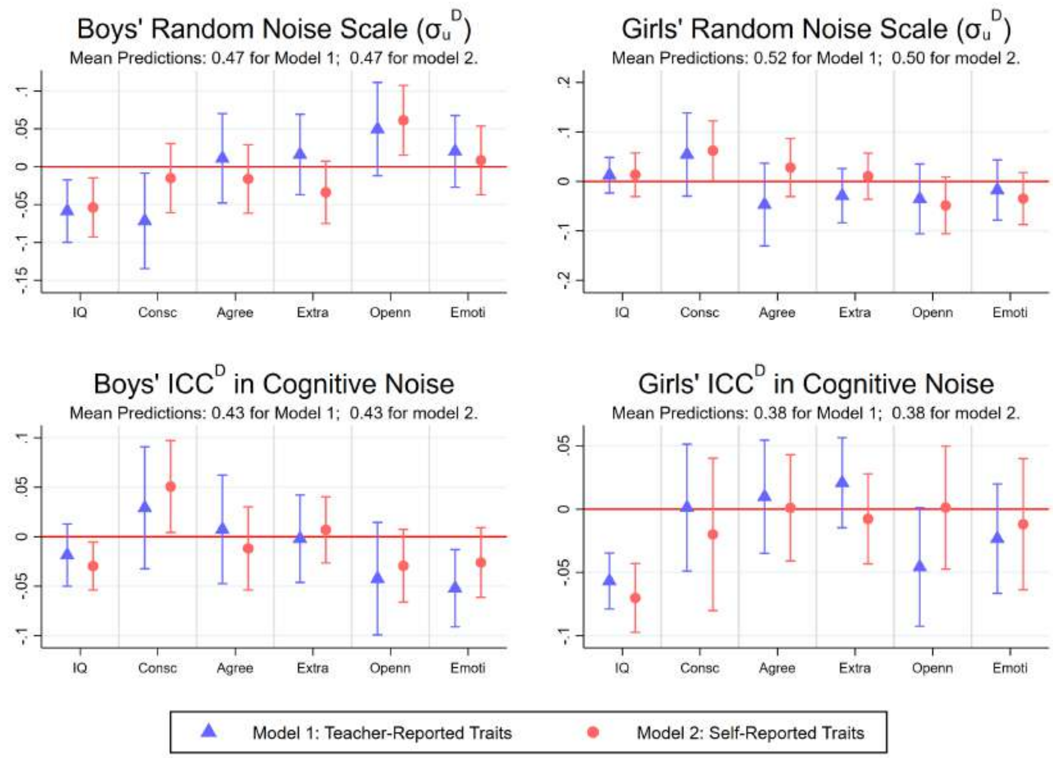


Figure D.6: Teacher vs. Self Reports: Transitory Noise σ_{u^D} and ICC^D (CRRA)

Notes: We compute predictions for each observation and average them over the sample. Blue triangles: Model 1 (teacher-reported traits, our baseline). Red circles: Model 2 (self-reported traits). Mean predictions appear in each panel's subtitle. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

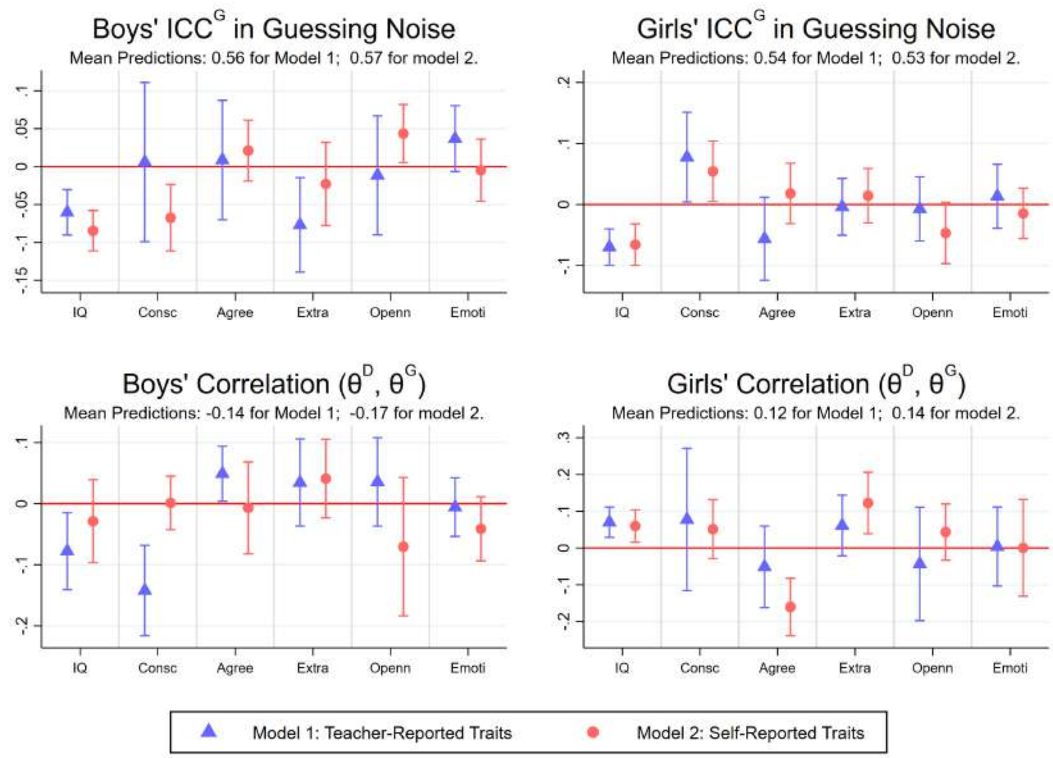


Figure D.7: Teacher vs. Self Reports: Guessing Persistence ICC^G and Deliberation–Guessing Correlation (CRRA)

Notes: We compute predictions for each observation and average them over the sample. Blue triangles: Model 1 (teacher-reported traits, our baseline). Red circles: Model 2 (self-reported traits). Mean predictions appear in each panel's subtitle. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

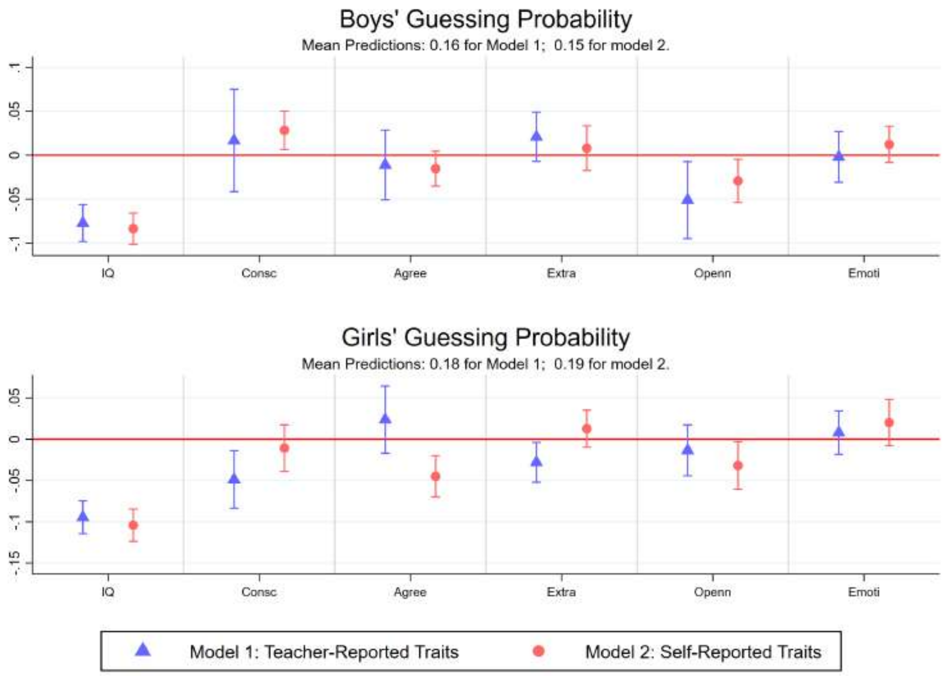


Figure D.8: Teacher vs. Self Reports: Guessing Probability (CRRA)

Notes: We compute predictions for each observation and average them over the sample. Blue triangles: Model 1 (teacher-reported traits, our baseline). Red circles: Model 2 (self-reported traits). Mean predictions appear in each panel's subtitle. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

E Descriptive Statistics Tables

Table E.1: Descriptive Statistics: Grade 5 in 2019 (Full Sample)

	Boys		**Girls**		**Difference**	
	Mean	S.E.	Mean	S.E.	Mean	S.E.
IQ and personality traits						
IQ Index (Standardized)[a]	−0.012	0.031	0.012	0.029	0.024	0.043
Conscientiousness	0.246	0.021	0.631	0.020	0.385***	0.029
Agreeableness	0.581	0.026	0.837	0.024	0.256***	0.036
Extraversion	0.066	0.023	0.064	0.025	−0.002	0.034
Openness to Experiences	0.437	0.020	0.454	0.020	0.017	0.028
Emotional Stability[b]	0.282	0.022	0.461	0.021	0.179***	0.031
Demographics						
Age	10.707	0.010	10.655	0.010	−0.052***	0.014
Living in Urban Area	0.227	0.013	0.233	0.013	0.006	0.018
Parental Education						
Father Went to High School	0.305	0.014	0.348	0.014	0.043**	0.020
Mother Went to High School	0.249	0.013	0.288	0.014	0.039**	0.019
Family Structure and Migration Status						
Having Any Sibling	0.366	0.014	0.413	0.015	0.047**	0.021
Left-Behind by Father Only	0.250	0.010	0.253	0.010	0.003	0.014
Left-Behind by Mother Only	0.069	0.006	0.056	0.005	−0.013*	0.007
Left-Behind by Both	0.170	0.009	0.154	0.009	−0.017	0.013
LBC Volatility Index	1.454	0.053	1.423	0.052	−0.031	0.074
Observations	1,115		1,089		2,204	

Notes: Full descriptive statistics for Grade 5 students surveyed in 2019. Columns report means with standard errors (S.E.) in the adjacent column. The Difference column reports the girls-minus-boys mean difference, with significance from a two-sample test of equality of means. The standardized IQ index is estimated using an IRT-3PL model and standardized to mean 0 and standard deviation 1 in the pooled sample of boys and girls. Big Five traits are predicted using the factor model in Heckman et al. (2013), with estimation-error corrections via the Bartlett GLS estimator. Binary variables are coded as 0/1 indicators with means representing proportions. Left-behind status variables represent averages of indicator variables across nine semesters (Grade 1, Semester 1 through Grade 5, Semester 1). The LBC Volatility Index measures the number of transitions in migration status during these nine semesters. [a]Standardized relative to the pooled sample of boys and girls. [b]When reverse-coded, also known as Neuroticism. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table E.2: Correlation Matrix: Grade 5 in 2019 - Full Sample

	IQ	Consc.	Agree.	Extra.	Openn.	Emoti.
IQ	1.000					
Conscientiousness	0.197	1.000				
Agreeableness	0.146	0.640	1.000			
Extraversion	0.138	0.380	0.321	1.000		
Openness	0.248	0.484	0.423	0.645	1.000	
Emotional Stability	0.129	0.559	0.534	0.469	0.475	1.000

Notes: This table presents pairwise correlations between IQ (IQ Index) and Big Five personality traits for all Grade 5 students in 2019 (N = 2,204). Personality traits follow the measurement method in Heckman et al. (2013). Consc. = Conscientiousness. Agree. = Agreeableness. Extra. = Extraversion. Openn. = Openness to Experiences. Emoti. = Emotional Stability.

Table E.3: Correlation Matrix: Grade 5 in 2019 - Boys

	IQ	Consc.	Agree.	Extra.	Openn.	Emoti.
IQ	1.000					
Conscientiousness	0.204	1.000				
Agreeableness	0.159	0.577	1.000			
Extraversion	0.141	0.389	0.313	1.000		
Openness	0.266	0.491	0.420	0.630	1.000	
Emotional Stability	0.141	0.528	0.489	0.436	0.439	1.000

Notes: This table presents pairwise correlations between IQ (IQ Index) and Big Five personality traits for Grade 5 boys in 2019 (N = 1,115). Personality traits follow the measurement method in Heckman et al. (2013). Consc. = Conscientiousness. Agree. = Agreeableness. Extra. = Extraversion. Openn. = Openness to Experiences. Emoti. = Emotional Stability.

Table E.4: Correlation Matrix: Grade 5 in 2019 - Girls

	IQ	Consc.	Agree.	Extra.	Openn.	Emoti.
IQ	1.000					
Conscientiousness	0.199	1.000				
Agreeableness	0.130	0.692	1.000			
Extraversion	0.135	0.405	0.339	1.000		
Openness	0.227	0.509	0.433	0.662	1.000	
Emotional Stability	0.113	0.576	0.568	0.513	0.521	1.000

Notes: This table presents pairwise correlations between IQ (IQ Index) and Big Five personality traits for Grade 5 girls in 2019 ($N = 1{,}089$). Personality traits follow the measurement method in Heckman et al. (2013). Consc. = Conscientiousness. Agree. = Agreeableness. Extra. = Extraversion. Openn. = Openness to Experiences. Emoti. = Emotional Stability.

Table E.5: Determinants of IQ and Personality Traits: Boys

Boys in Grade 5 in 2019	Personality Traits Evaluated by Teachers, 20 Items					
	IQ (1)	Consc. (2)	Agree. (3)	Extra. (4)	Openn. (5)	Stability (6)
Living in Urban Area	0.176^{**}	-0.112^{**}	0.142^{**}	-0.073	-0.064	-0.225^{***}
	(0.079)	(0.049)	(0.060)	(0.054)	(0.051)	(0.054)
Father Went to High School	0.094	0.117^{**}	0.119^{**}	0.030	0.105^{**}	0.082
	(0.069)	(0.048)	(0.059)	(0.054)	(0.048)	(0.053)
Mother Went to High School	0.043	0.041	-0.061	0.077	0.033	-0.032
	(0.075)	(0.051)	(0.062)	(0.058)	(0.051)	(0.058)
Having Siblings	-0.025	-0.050	-0.090^{*}	-0.060	-0.020	0.031
	(0.065)	(0.044)	(0.054)	(0.047)	(0.043)	(0.047)
LBC by Father	0.110	0.075	0.070	0.080	0.026	-0.013
	(0.099)	(0.067)	(0.085)	(0.073)	(0.064)	(0.074)
LBC by Mother	0.210	0.023	-0.077	0.141	0.065	0.053
	(0.152)	(0.122)	(0.158)	(0.140)	(0.118)	(0.149)
LBC by Both	-0.012	-0.037	-0.009	0.017	-0.032	-0.010
	(0.111)	(0.072)	(0.093)	(0.083)	(0.072)	(0.079)
LBC Volatility	-0.032^{*}	-0.018	-0.007	-0.046^{***}	-0.032^{***}	-0.016
	(0.019)	(0.013)	(0.015)	(0.014)	(0.012)	(0.014)
Number of Students	1115	1115	1115	1115	1115	1115

Notes: OLS regressions with IQ and Big Five personality traits as dependent variables for boys. All dependent variables are standardized. Personality traits are teacher-reported using 20-item inventories. LBC = Left-Behind Children (parent migrated for work). LBC Volatility measures number of changes in LBC status during these nine semesters. Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table E.6: Determinants of IQ and Personality Traits: Girls

Girls in Grade 5 in 2019	Personality Traits Evaluated by Teachers, 20 Items					
	IQ	Consc.	Agree.	Extra.	Openn.	Stability
	(1)	(2)	(3)	(4)	(5)	(6)
Living in Urban Area	0.070	-0.248^{***}	-0.067	-0.133^{**}	-0.183^{***}	-0.287^{***}
	(0.073)	(0.046)	(0.059)	(0.063)	(0.050)	(0.048)
Father Went to High School	0.161^{**}	0.143^{***}	0.097^{*}	0.122^{**}	0.127^{***}	0.102^{**}
	(0.067)	(0.045)	(0.057)	(0.058)	(0.044)	(0.049)
Mother Went to High School	0.111	0.107^{**}	0.118^{**}	0.090	0.145^{***}	0.081
	(0.071)	(0.045)	(0.058)	(0.059)	(0.046)	(0.050)
Having Siblings	-0.050	-0.024	0.002	0.028	-0.032	0.059
	(0.062)	(0.040)	(0.049)	(0.050)	(0.040)	(0.043)
LBC by Father	-0.002	-0.019	0.007	-0.027	-0.026	-0.002
	(0.095)	(0.061)	(0.078)	(0.084)	(0.067)	(0.072)
LBC by Mother	-0.197	-0.319^{**}	-0.145	-0.443^{**}	-0.338^{**}	-0.125
	(0.193)	(0.159)	(0.178)	(0.176)	(0.144)	(0.153)
LBC by Both	-0.106	-0.094	0.014	-0.003	0.006	0.006
	(0.116)	(0.071)	(0.091)	(0.092)	(0.071)	(0.076)
LBC Volatility	-0.033^{*}	-0.008	-0.007	0.006	0.000	-0.008
	(0.019)	(0.013)	(0.016)	(0.016)	(0.013)	(0.014)
Number of Students	1089	1089	1089	1089	1089	1089

Notes: OLS regressions with IQ and Big Five personality traits as dependent variables for girls. All dependent variables are standardized. Personality traits are teacher-reported using 20-item inventories. LBC = Left-Behind Children (parent migrated for work). LBC Volatility measures number of changes in LBC status during these nine semesters. Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table E.7: Intra-Class Correlation in Deliberation and Guessing (by Gender)

	ICC: Deliberation		ICC: Guessing	
	Boys	Girls	Boys	Girls
Expo-Power	0.43	0.38	0.57	0.53
CRRA	0.43	0.38	0.56	0.54

Notes: ICC $= \sigma_\theta^2/(\sigma_\theta^2 + \sigma_u^2)$, where σ_θ^2 is the variance of persistent individual components and σ_u^2 is the variance of transitory noise. Values shown are evaluated over the sampling distribution of covariate values, since σ_u^2 depends on $(\boldsymbol{X_i})$ as specified in equation (19). ICC^D measures the proportion of variance in deliberation noise attributable to persistent individual differences. ICC^G measures the proportion of variance in guessing behavior attributable to persistent individual differences. Higher values indicate greater stability of these behavioral patterns within individuals across tasks.

Table E.8: Reduced-Form Analysis: Fraction of MPLs with Inconsistent Choices

	Boys			Girls		
Grade 5 in 2019	(1)	(2)	(3)	(4)	(5)	(6)
Living in Urban Area	-0.036^{*}		-0.031	-0.048^{**}		-0.053^{**}
	(0.022)		(0.022)	(0.022)		(0.022)
Father Went to HS	0.016		0.024	-0.058^{***}		-0.046^{**}
	(0.020)		(0.020)	(0.021)		(0.020)
Mother Went to HS	-0.047^{**}		-0.045^{**}	0.001		0.009
	(0.021)		(0.021)	(0.022)		(0.021)
Having Any Sibling	-0.041^{**}		-0.040^{**}	-0.024		-0.025
	(0.018)		(0.018)	(0.018)		(0.018)
LBC by Father Only	0.022		0.025	-0.047		-0.048
	(0.027)		(0.026)	(0.030)		(0.030)
LBC by Mother Only	0.033		0.043	0.018		-0.008
	(0.047)		(0.046)	(0.058)		(0.056)
LBC by Both	0.030		0.027	-0.033		-0.041
	(0.032)		(0.031)	(0.034)		(0.033)
LBC Volatility Index	0.011^{**}		0.009^{*}	0.007		0.006
	(0.005)		(0.005)	(0.006)		(0.006)
IQ		-0.049^{***}	-0.047^{***}		-0.051^{***}	-0.047^{***}
		(0.010)	(0.010)		(0.010)	(0.010)
Conscientiousness		0.007	0.005		-0.031	-0.034^{*}
		(0.017)	(0.018)		(0.019)	(0.019)
Agreeableness		0.014	0.016		0.006	0.011
		(0.013)	(0.013)		(0.015)	(0.015)
Extraversion		0.034^{**}	0.036^{**}		-0.036^{**}	-0.033^{**}
		(0.015)	(0.015)		(0.015)	(0.015)
Openness		-0.069^{***}	-0.068^{***}		0.022	0.021
		(0.018)	(0.017)		(0.020)	(0.020)
Emotional Stability		0.006	0.002		0.006	0.001
		(0.014)	(0.015)		(0.018)	(0.018)
R-Square	0.022	0.050	0.069	0.021	0.042	0.058
Observations	1115	1115	1115	1089	1089	1089

Notes: Linear probability models with dependent variable equal to the fraction of multiple price lists (MPLs) where the child switched back to a lower-variance option after choosing a higher-variance one, even when the previous task is less favorable to the higher-variance option in terms of either rewards or winning probability. Inconsistent behavior may indicate confusion, inattention, or inconsistent preferences. Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Columns (1) and (4) include only demographic and family structure variables. Columns (2) and (5) include only cognitive and personality traits. Columns (3) and (6) include all covariates.

F Socioeconomic Status vs. IQ and Personality Traits

Tables E.5 and E.6 demonstrate that background variables predict IQ and Big Five traits. A natural issue is whether background variables have direct effects on preferences. We address this question by comparing three specifications: Model 1 includes only background variables (urbanicity, parental education, sibling structure, and left-behind status), Model 2 includes only IQ and Big Five personality traits, and Model 3 includes both sets of covariates. All models incorporate the full ICC structure and guessing behavior. Figures F.1–F.7 present results for our preferred Expo-Power specifications.[26] Predictions are broadly similar across the three models under both functional forms, with the mean predictions shown in each panel's subtitle.

Under Expo-Power, Model 1 (background only) yields $r = 0.83$ and $\beta = 0.07$, whereas Models 2 and 3 yield $r = 0.76$–0.77 and $\beta = 0.09$. The risk aversion curvature falls modestly and wealth-dependence rises slightly when traits are added, leaving overall risk aversion broadly stable. For girls, both $r = 0.58$ and $\beta = 0.09$ are invariant across the three sets of covariates. This suggests that background variables play some role in boys' preference structure. IQ and personality traits play different roles in the level and slope components. For girls the decomposition is robust to choice of conditioning variables.

The same pattern holds for deliberation noise parameters. Under Expo-Power, transitory noise σ_{u^D} shows predictions of 0.51–0.58 for boys and 0.86–0.92 for girls across models. ICC^D remains at 0.42–0.44 for boys and 0.38 for girls under Expo-Power.[27]

For ICC^G, boys show a modest reduction when adding IQ and traits (0.64 in Model 1 to 0.57 in Models 2 and 3), reflecting the genuine predictive power of personality traits—particularly Emotional Stability—for persistent guessing patterns. Girls' ICC^G shows a smaller reduction from 0.57 in Model 1 to 0.53–0.54 in Models 2 and 3.

The correlation between deliberation and guessing random effects $\rho(\theta^D, \theta^G)$ provides

[26] Figures F.8–F.13 present results for CRRA.

[27] Under CRRA the results are similar.

perhaps the most striking evidence. The gender difference in correlation patterns—negative for boys ($\rho = -0.15$ to -0.22) and positive for girls ($\rho = 0.06$ to 0.11)—persists across all three specifications under Expo-Power. The opposite-signed correlations reflect genuine differences by gender in cognitive architecture rather than some confound with socioeconomic background or measured traits.

Guessing probability is estimated to be 0.12–0.15 for boys and 0.17–0.19 for girls across models under Expo-Power, stable within 3 percentage points.

Expo-Power Specification

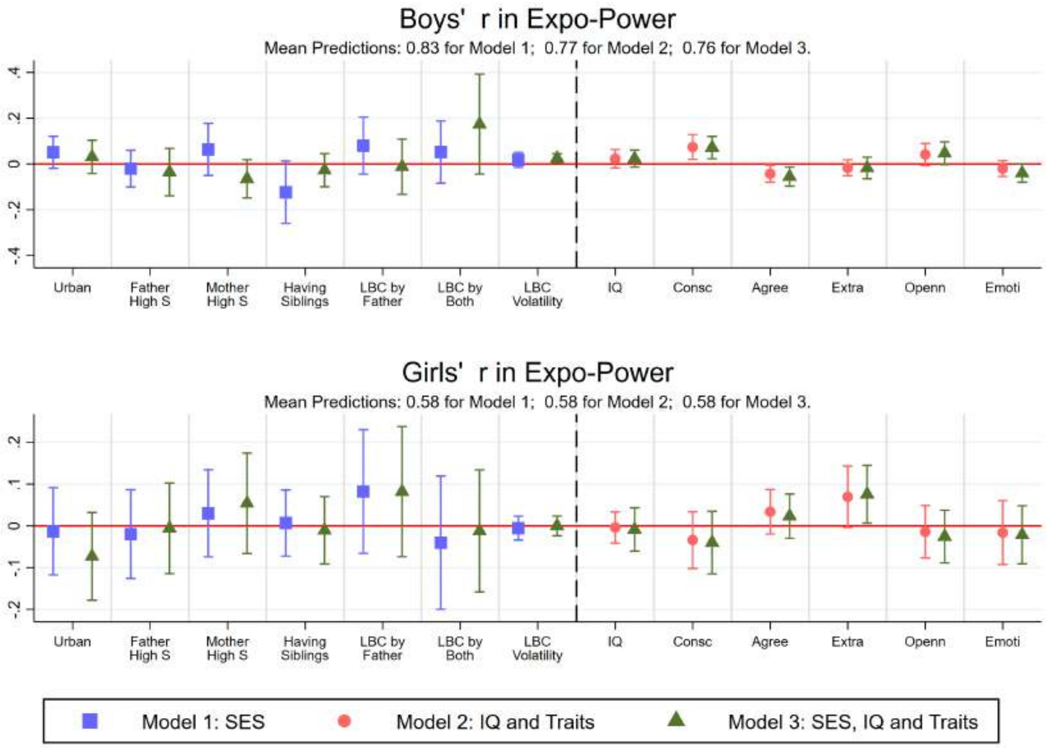


Figure F.1: SES vs. Traits: Risk Aversion Curvature r under Expo-Power

Notes: We compute predictions for each observation and average them over the sample. For boys, the curvature shift between Model 1 and Models 2–3 is offset by movements in β (Figure F.2), leaving overall risk aversion stable, while for girls r is unchanged across specifications. Blue squares: Model 1 (SES only). Red circles: Model 2 (IQ and Big Five traits only). Green triangles: Model 3 (both). Mean predictions across the three specifications appear in each panel's subtitle. Top panel: boys, bottom panel: girls. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

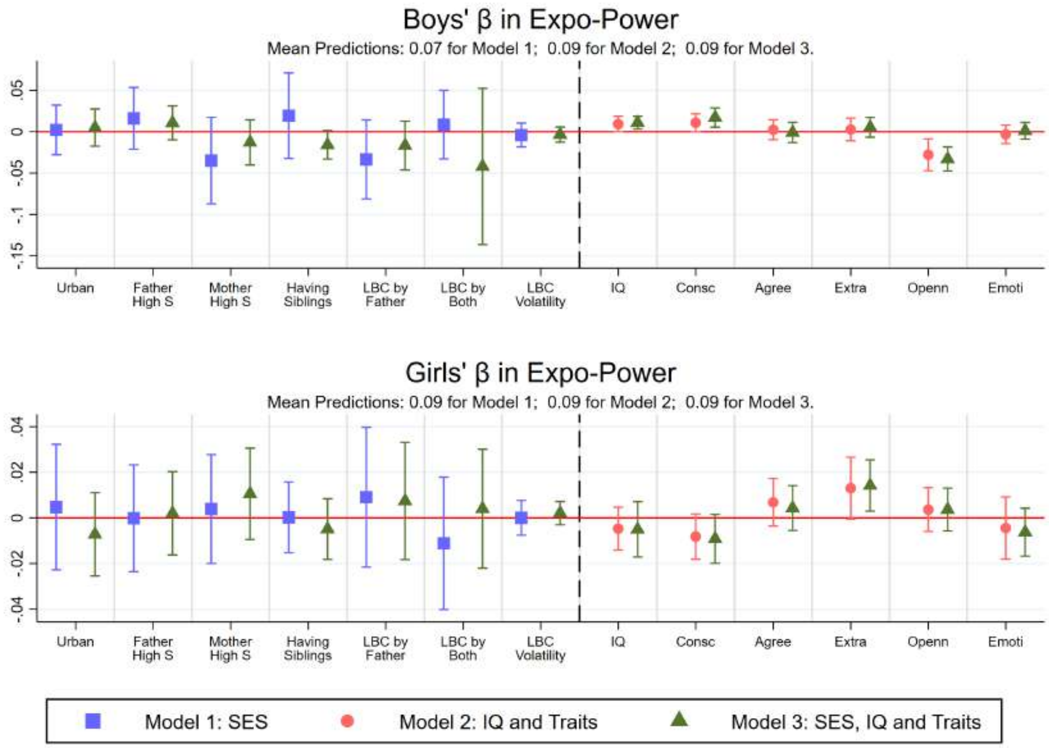


Figure F.2: SES vs. Traits: Wealth-Dependence β under Expo-Power

Notes: We compute predictions for each observation and average them over the sample. For boys, SES alone yields a lower β, while adding IQ and traits raises it with an offsetting reduction in r (Figure F.1), while for girls β is stable. Blue squares: Model 1 (SES only). Red circles: Model 2 (IQ and Big Five traits only). Green triangles: Model 3 (both). Mean predictions across the three specifications appear in each panel's subtitle. Top panel: boys, bottom panel: girls. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

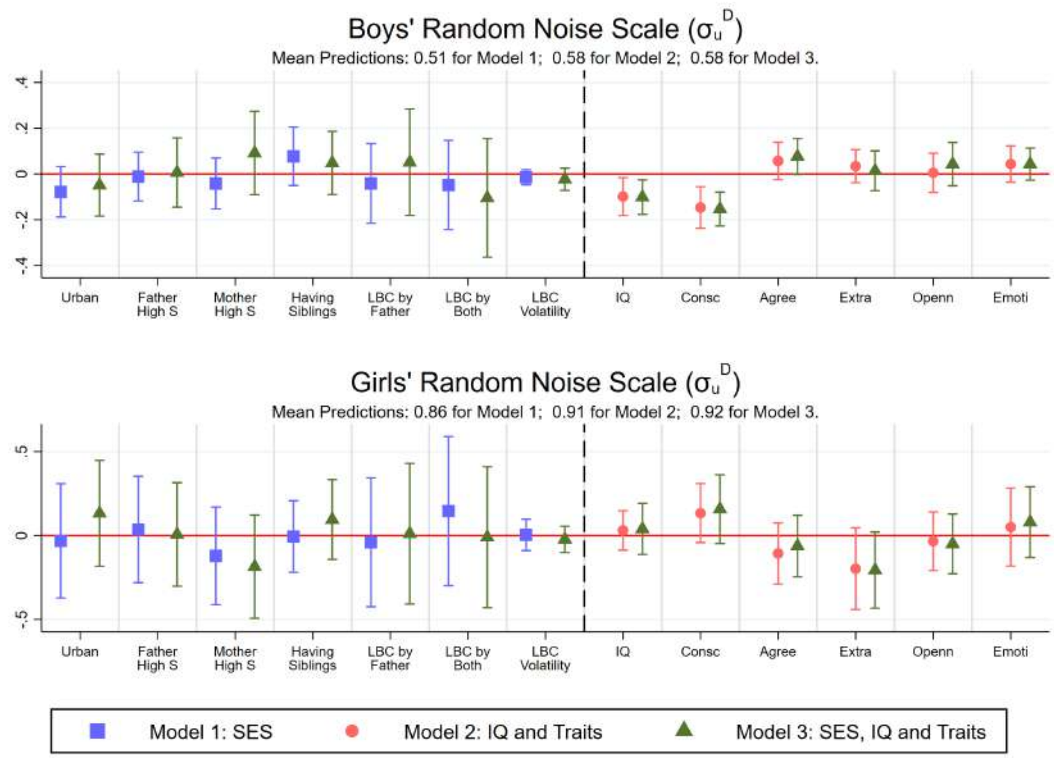


Figure F.3: SES vs. Traits: Transitory Noise σ_{u^D} under Expo-Power

Notes: We compute predictions for each observation and average them over the sample. Blue squares: Model 1 (SES only). Red circles: Model 2 (IQ and Big Five traits only). Green triangles: Model 3 (both). Mean predictions across the three specifications appear in each panel's subtitle. Top panel: boys, bottom panel: girls. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

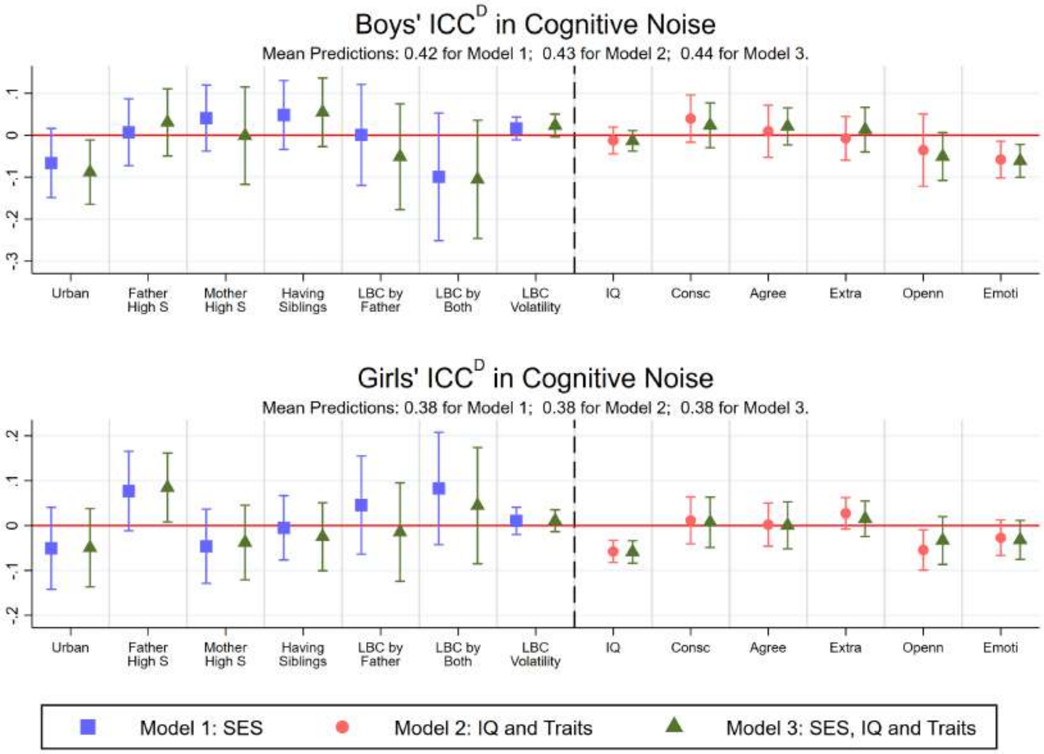


Figure F.4: SES vs. Traits: ICC^D under Expo-Power

Notes: We compute predictions for each observation and average them over the sample. Blue squares: Model 1 (SES only). Red circles: Model 2 (IQ and Big Five traits only). Green triangles: Model 3 (both). Mean predictions across the three specifications appear in each panel's subtitle. Top panel: boys, bottom panel: girls. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

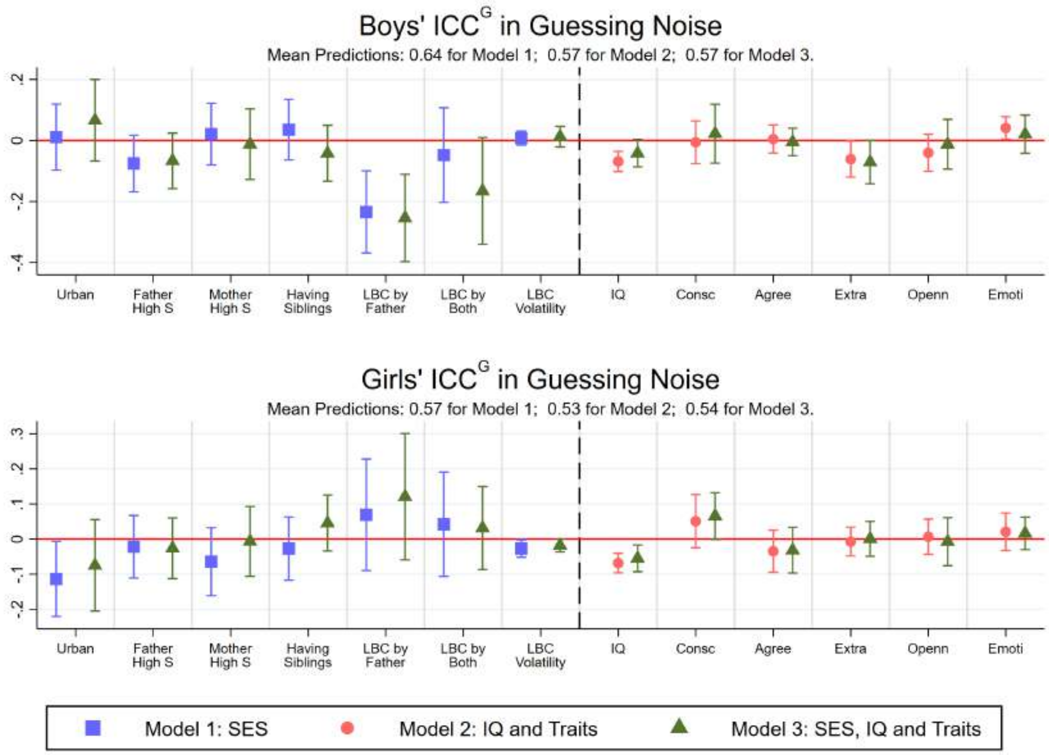


Figure F.5: SES vs. Traits: ICC^{G} under Expo-Power

Notes: We compute predictions for each observation and average them over the sample. Boys show a modest reduction in ICC^{G} when IQ and traits are added, while girls are stable. Blue squares: Model 1 (SES only). Red circles: Model 2 (IQ and Big Five traits only). Green triangles: Model 3 (both). Mean predictions across the three specifications appear in each panel's subtitle. Top panel: boys, bottom panel: girls. Error bars show 90% confidence intervals. N = 1,115 boys, N = 1,089 girls.

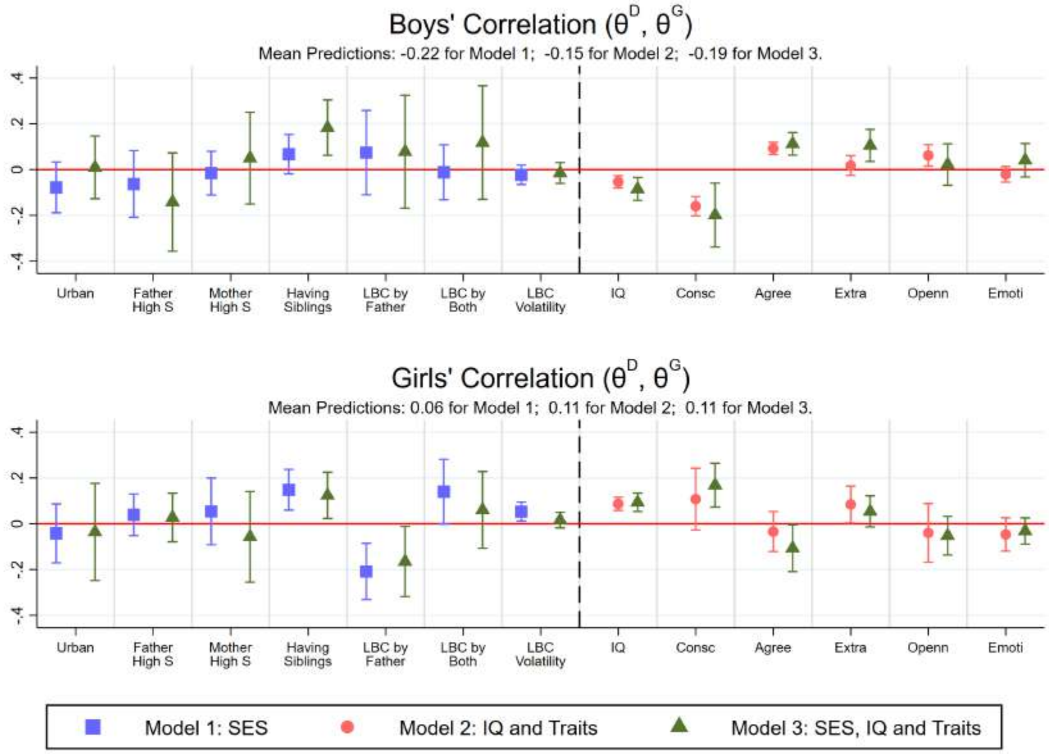


Figure F.6: SES vs. Traits: Correlation $\rho(\theta^D, \theta^G)$ under Expo-Power

Notes: We compute predictions for each observation and average them over the sample. The opposite-signed correlations by gender (negative for boys, positive for girls) persist across all specifications. Blue squares: Model 1 (SES only). Red circles: Model 2 (IQ and Big Five traits only). Green triangles: Model 3 (both). Mean predictions across the three specifications appear in each panel's subtitle. Top panel: boys, bottom panel: girls. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

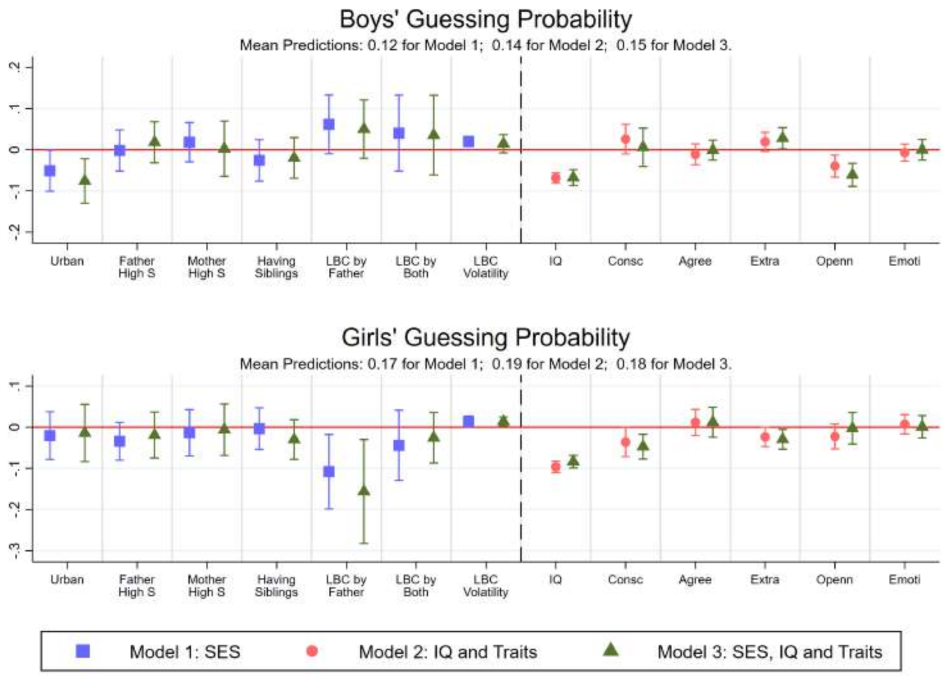


Figure F.7: SES vs. Traits: Guessing Probability under Expo-Power

Notes: We compute predictions for each observation and average them over the sample. Blue squares: Model 1 (SES only). Red circles: Model 2 (IQ and Big Five traits only). Green triangles: Model 3 (both). Mean predictions across the three specifications appear in each panel's subtitle. Top panel: boys, bottom panel: girls. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

CRRA Specification

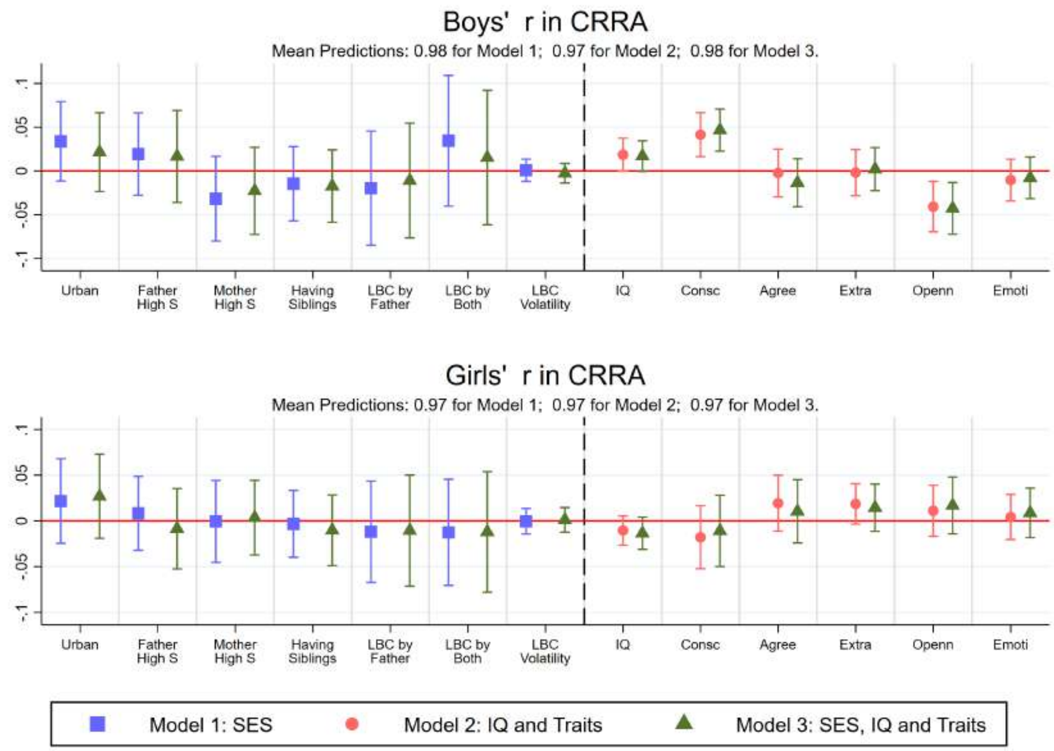


Figure F.8: SES vs. Traits: Risk Aversion r under CRRA

Notes: We compute predictions for each observation and average them over the sample. Blue squares: Model 1 (SES only). Red circles: Model 2 (IQ and Big Five traits only). Green triangles: Model 3 (both). Mean predictions across the three specifications appear in each panel's subtitle. Top panel: boys, bottom panel: girls. Error bars show 90% confidence intervals. N = 1,115 boys, N = 1,089 girls.

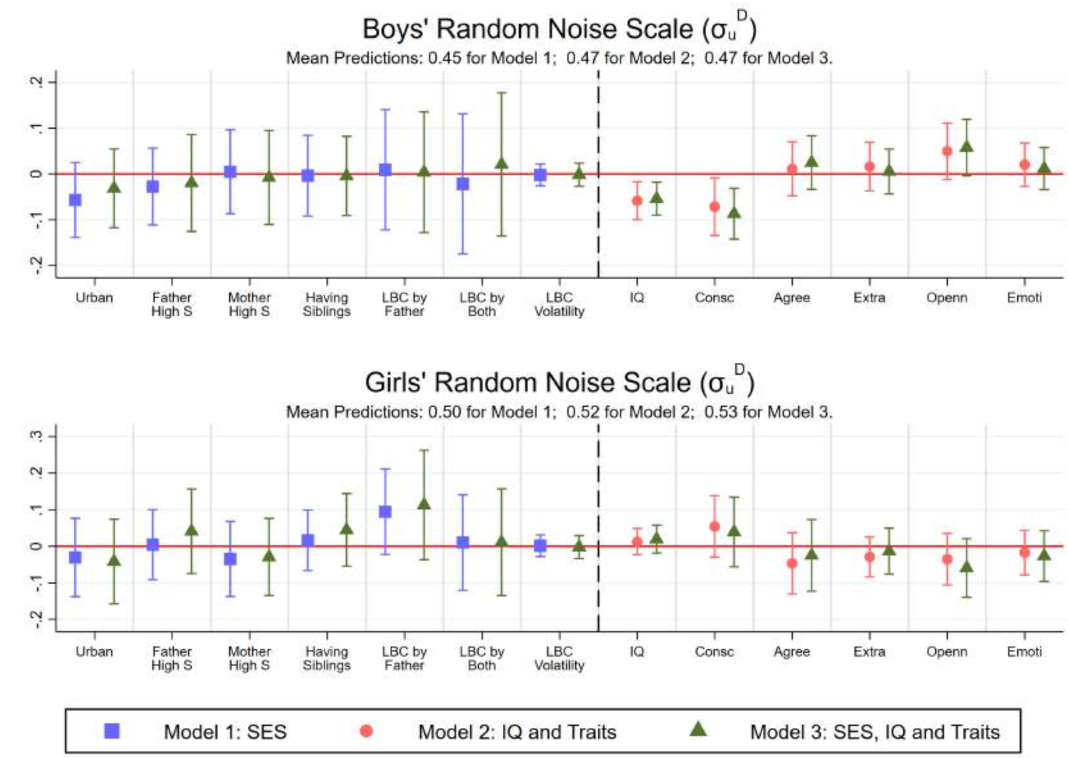


Figure F.9: SES vs. Traits: Transitory Noise σ_{u^D} under CRRA

Notes: We compute predictions for each observation and average them over the sample. Blue squares: Model 1 (SES only). Red circles: Model 2 (IQ and Big Five traits only). Green triangles: Model 3 (both). Mean predictions across the three specifications appear in each panel's subtitle. Top panel: boys, bottom panel: girls. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

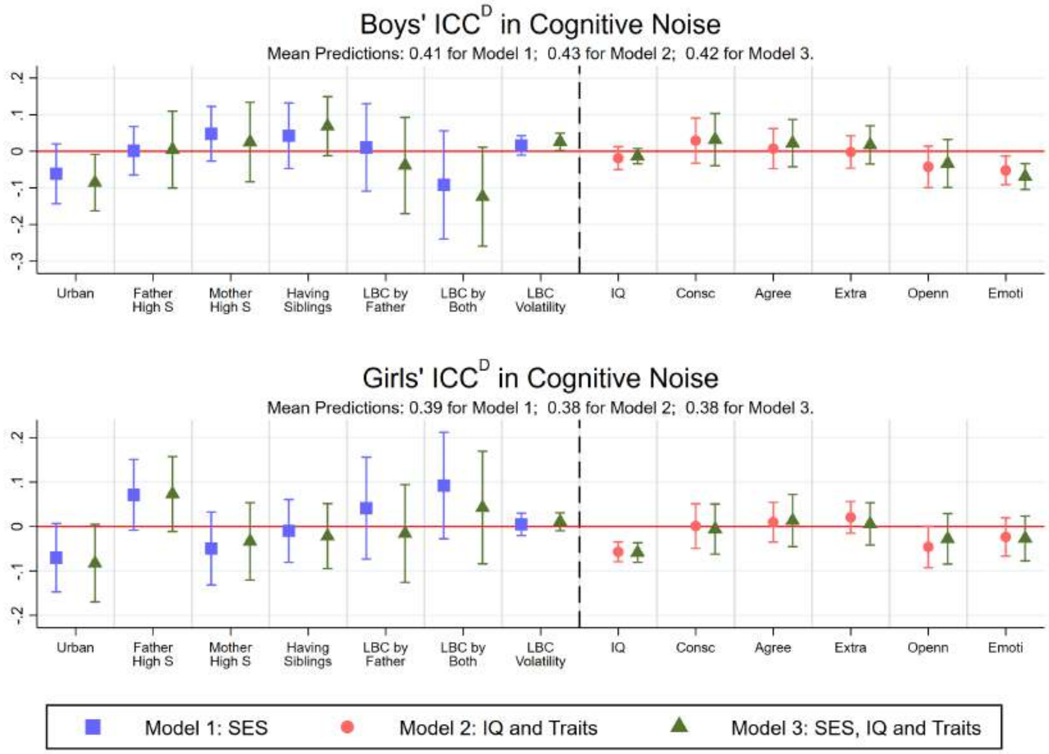


Figure F.10: SES vs. Traits: ICC^D under CRRA

Notes: We compute predictions for each observation and average them over the sample. Blue squares: Model 1 (SES only). Red circles: Model 2 (IQ and Big Five traits only). Green triangles: Model 3 (both). Mean predictions across the three specifications appear in each panel's subtitle. Top panel: boys, bottom panel: girls. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

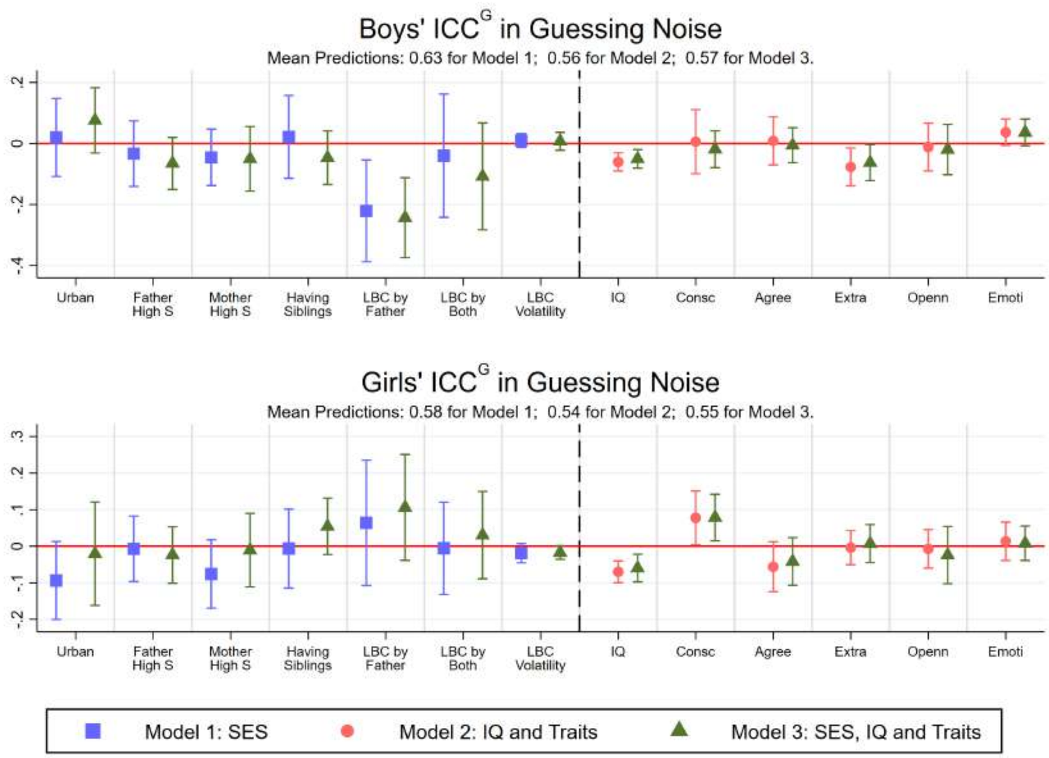


Figure F.11: SES vs. Traits: ICC^G under CRRA

Notes: We compute predictions for each observation and average them over the sample. Blue squares: Model 1 (SES only). Red circles: Model 2 (IQ and Big Five traits only). Green triangles: Model 3 (both). Mean predictions across the three specifications appear in each panel's subtitle. Top panel: boys, bottom panel: girls. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

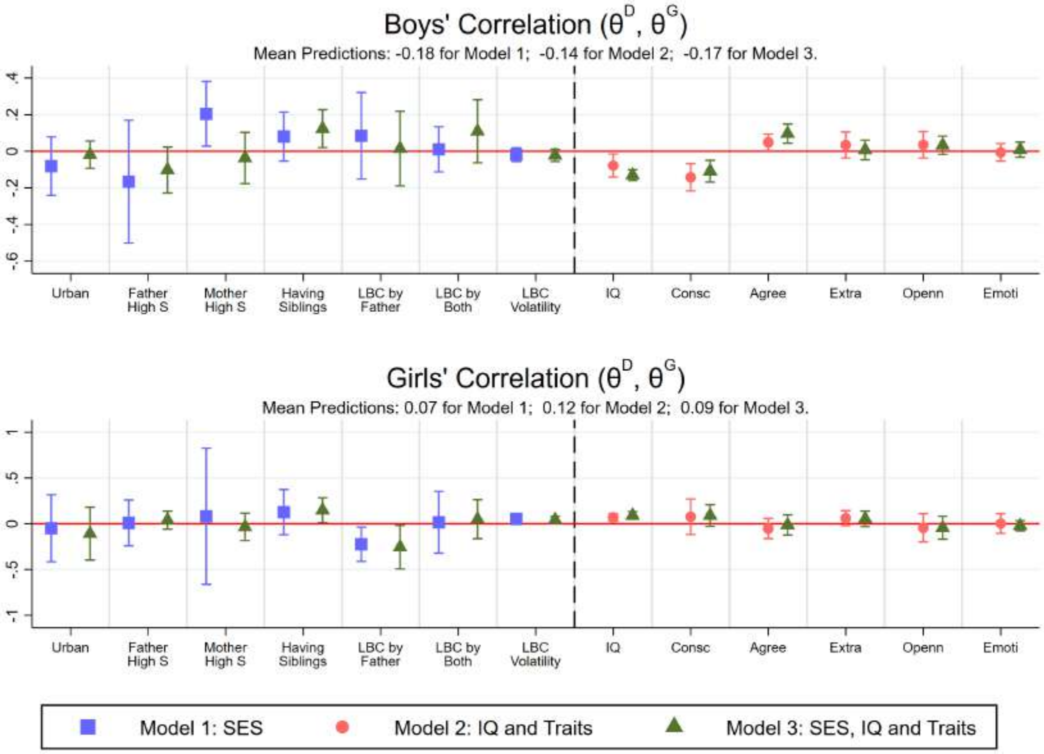


Figure F.12: SES vs. Traits: Correlation $\rho(\theta^D, \theta^G)$ under CRRA

Notes: We compute predictions for each observation and average them over the sample. Blue squares: Model 1 (SES only). Red circles: Model 2 (IQ and Big Five traits only). Green triangles: Model 3 (both). Mean predictions across the three specifications appear in each panel's subtitle. Top panel: boys, bottom panel: girls. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

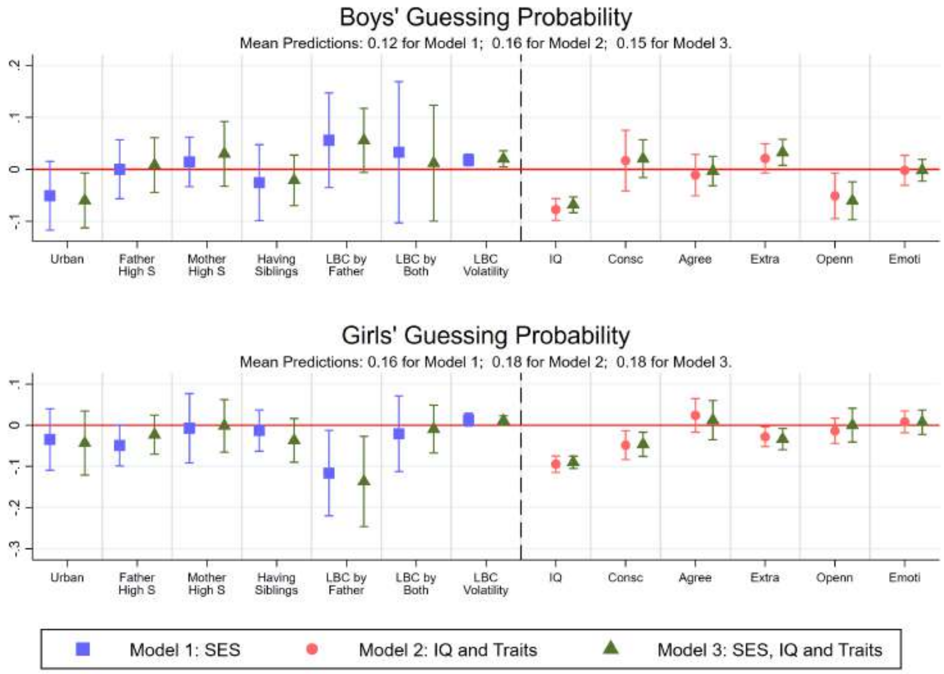


Figure F.13: SES vs. Traits: Guessing Probability under CRRA

Notes: We compute predictions for each observation and average them over the sample. Blue squares: Model 1 (SES only). Red circles: Model 2 (IQ and Big Five traits only). Green triangles: Model 3 (both). Mean predictions across the three specifications appear in each panel's subtitle. Top panel: boys, bottom panel: girls. Error bars show 90% confidence intervals. N = 1,115 boys, N = 1,089 girls.

G Sensitivity of Estimates to the Assumed Reference Wealth (w)

This appendix examines the sensitivity of our estimates to the assumed reference wealth w—the wealth level an individual retains when the game yields nothing. In our main analyses, this level is fixed at $w = 1$. Here we relax this assumption to examine whether our estimates are sensitive to it.

We re-estimate both CRRA and Expo-Power models fixing w at values around 1—

specifically, 0.2, 0.5, 1, 2, and 5. For each assumed level, total resources become $s + w$, where s is the payoff from the experimental game. This affects how relative risk aversion RRA(s+w) is evaluated and how the risk aversion curvatures translate into choice behavior.

We perform this sensitivity check for both the Expo-Power (Figures G.1–G.4) and CRRA (Figures G.5–G.8) utility functions. Each figure displays coefficient estimates across the five reference wealth levels for boys and girls, with the mean predictions at each level shown in the panel subtitles.

The marginal effects of IQ and personality traits on the estimated parameters shift with the assumed reference wealth w for both boys and girls under CRRA and Expo-Power. The main findings regarding gender differences in predicted risk attitudes at the sample mean remain unchanged: CRRA without ICC reproduces the conventional finding that girls are more risk averse than boys, CRRA with ICC narrows this gender gap, and Expo-Power with ICC reverses it. The role of wealth needs further exploration (see Chávez et al. (2026)).

The mean predictions reported in the panel subtitles separate cleanly into two groups. The parameters that carry the units of consumption respond to w, exactly as the identification implies, while the parameters that describe the structure of deliberation noise do not. Under Expo-Power, risk aversion curvature r rises monotonically with the assumed reference wealth, from 0.66 at $w = 0.2$ to 0.77 at the baseline and on to 0.89 at $w = 2$ and 1.10 at $w = 5$ for boys, and from 0.52 to 0.58 to 0.89 for girls (Figure G.1). The transitory noise σ_{u^D} moves in the opposite direction and by even more, falling monotonically from 0.78 to 0.21 for boys and from 1.13 to 0.32 for girls across the same range, a drop of roughly three quarters (Figure G.2). Both movements are mechanical. A larger reference wealth rescales the argument $(s + w)$ of the utility function, so a higher r and a smaller transitory noise are needed to rationalize the same choices. The wealth-dependence β stays near 0.09 for boys across the whole range and rises modestly for girls, from 0.08 to 0.16.

By contrast, the parameters that describe the structure of deliberation noise are dimensionless and are essentially flat across all five reference wealth levels. Under Expo-Power, ICC^D holds near 0.43 for boys and 0.38 for girls at every w, ICC^G near 0.57 and 0.53, the deliberation–guessing correlation at -0.15 to -0.16 for boys and $+0.11$ to $+0.12$ for girls, and the guessing probability at 0.14–0.15 for boys and 0.18–0.19 for girls (Figures G.3–G.4). None of these moves by more than a point or two in the second decimal as w ranges over a factor of twenty-five.

The CRRA panels tell the same story. Risk aversion r rises from 0.91 to 1.16 for boys and from 0.91 to 1.15 for girls as w goes from 0.2 to 5, and σ_{u^D} falls from about 0.61 to 0.21 for boys and from 0.66 to 0.22 for girls (Figures G.5–G.6). The structural noise parameters are again invariant: ICC^D at 0.42–0.44 for boys and 0.37–0.39 for girls, ICC^G near 0.56–0.57 and 0.52–0.54, the correlation at about -0.14 for boys and $+0.12$ for girls, and the guessing probability at 0.15–0.16 and 0.18–0.19 (Figures G.7–G.8).

Two features of the gender comparison survive the entire sweep. The deliberation–guessing correlation is negative for boys and positive for girls at every w in both utility functions, and girls guess more often than boys at every w. The level of risk aversion is sensitive to the reference-wealth normalization, but the decision-noise architecture that separates boys from girls is not.

G.1 Expo-Power Specification

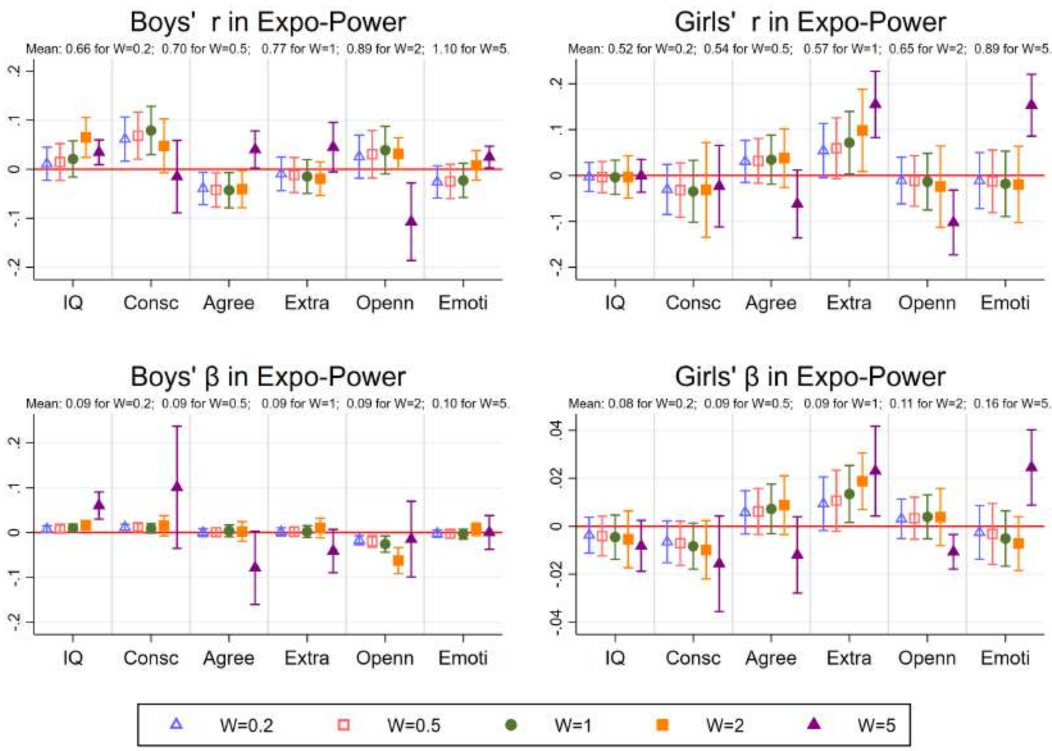


Figure G.1: Risk Aversion Parameters r and β Across Reference Wealth Levels (Expo-Power)

Notes: We compute predictions for each observation and average them over the sample. Top panels: risk aversion curvature r. Bottom panels: wealth-dependence β. The level of risk aversion rises mechanically with w while the structural pattern is preserved. Markers denote reference wealth levels: $w = 0.2$ (blue triangle), 0.5 (red square), 1 (green circle, baseline), 2 (orange square), 5 (purple triangle). Mean predictions at each w appear in the panel subtitles. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

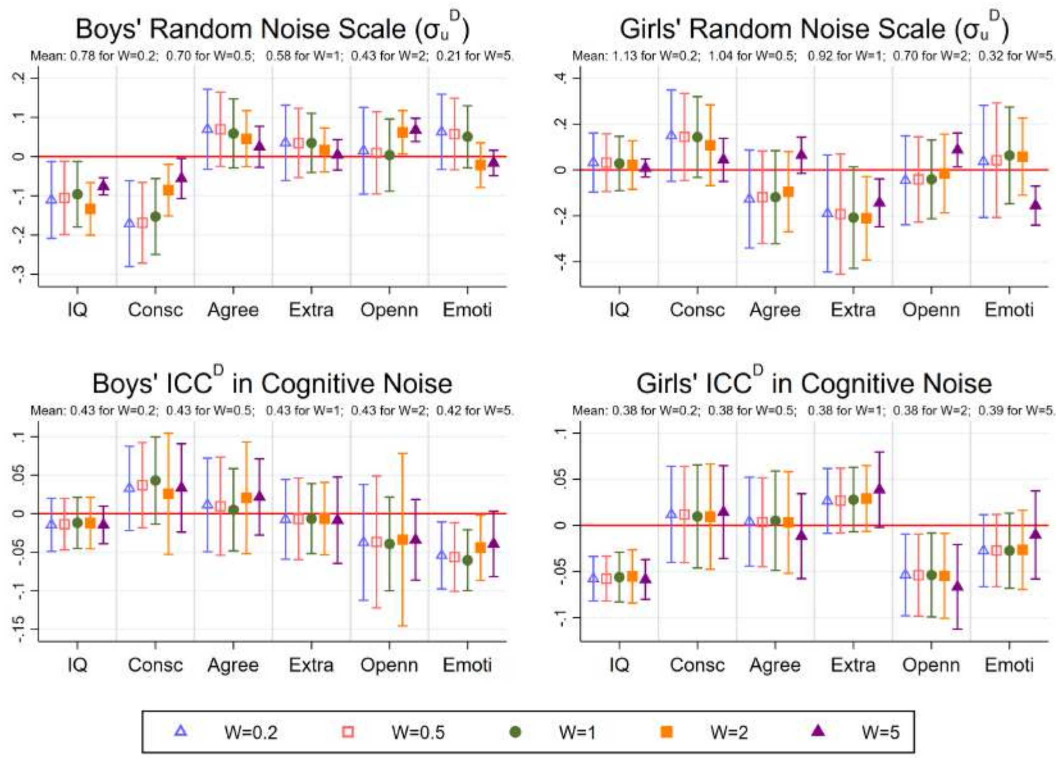


Figure G.2: Transitory Noise σ_{u^D} and ICC^D Across Reference Wealth Levels (Expo-Power)

Notes: We compute predictions for each observation and average them over the sample. Markers denote reference wealth levels: $w = 0.2$ (blue triangle), 0.5 (red square), 1 (green circle, baseline), 2 (orange square), 5 (purple triangle). Mean predictions at each w appear in the panel subtitles. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

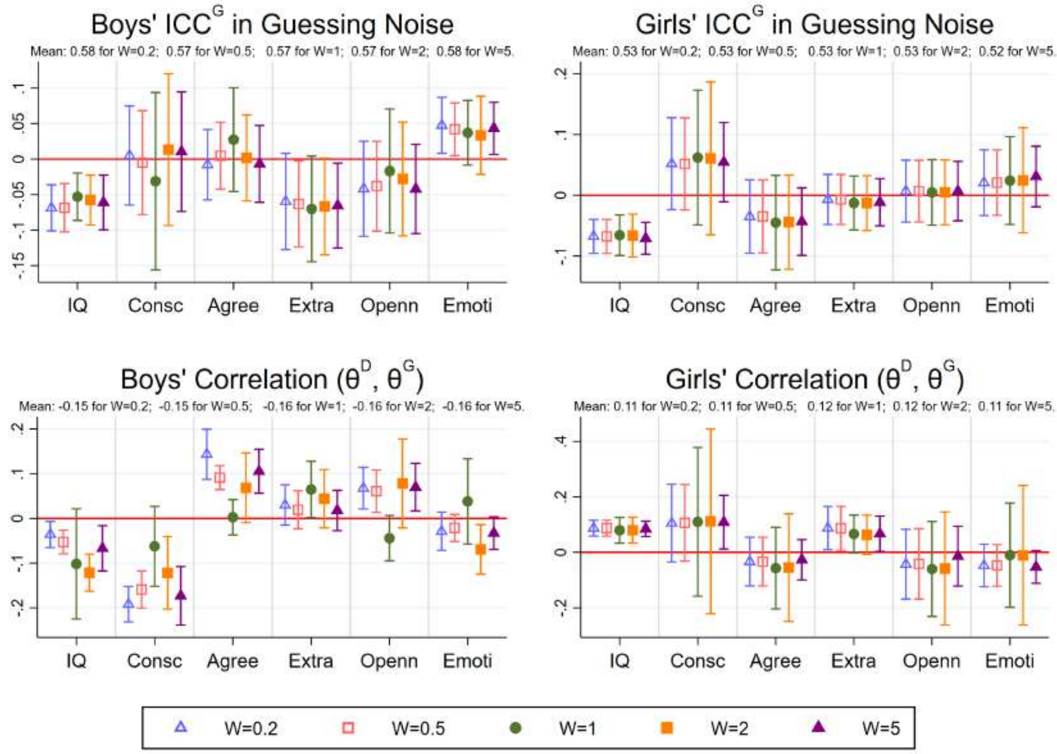


Figure G.3: Guessing Persistence ICC^G and Deliberation–Guessing Correlation Across Reference Wealth Levels (Expo-Power)

Notes: We compute predictions for each observation and average them over the sample. Markers denote reference wealth levels: $w = 0.2$ (blue triangle), 0.5 (red square), 1 (green circle, baseline), 2 (orange square), 5 (purple triangle). Mean predictions at each w appear in the panel subtitles. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

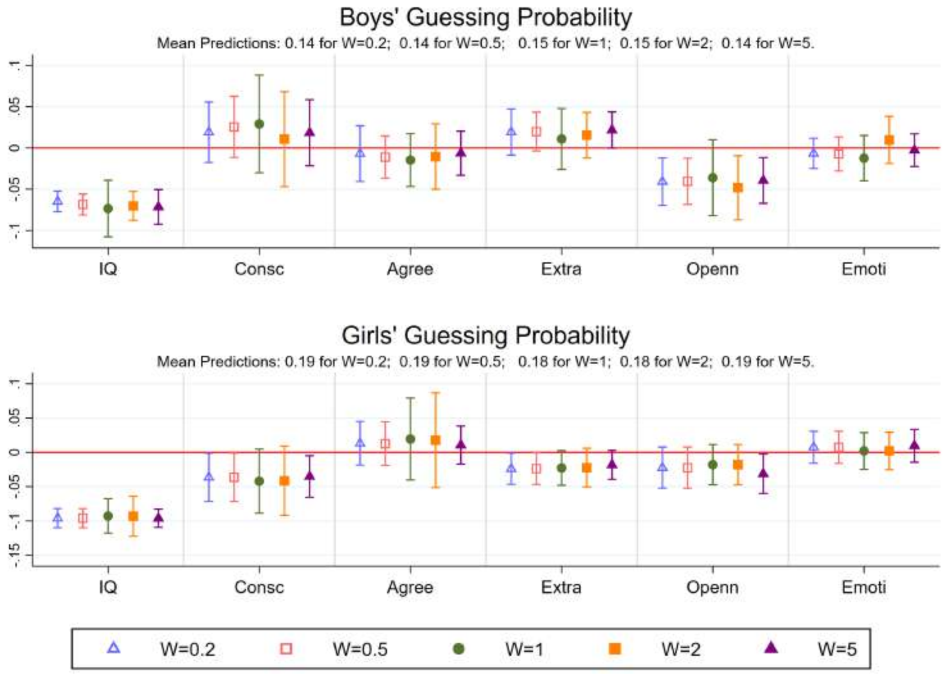


Figure G.4: Guessing Probability Across Reference Wealth Levels (Expo-Power)

Notes: We compute predictions for each observation and average them over the sample. Markers denote reference wealth levels: $w = 0.2$ (blue triangle), 0.5 (red square), 1 (green circle, baseline), 2 (orange square), 5 (purple triangle). Mean predictions at each w appear in the panel subtitles. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

G.2 CRRA Specification

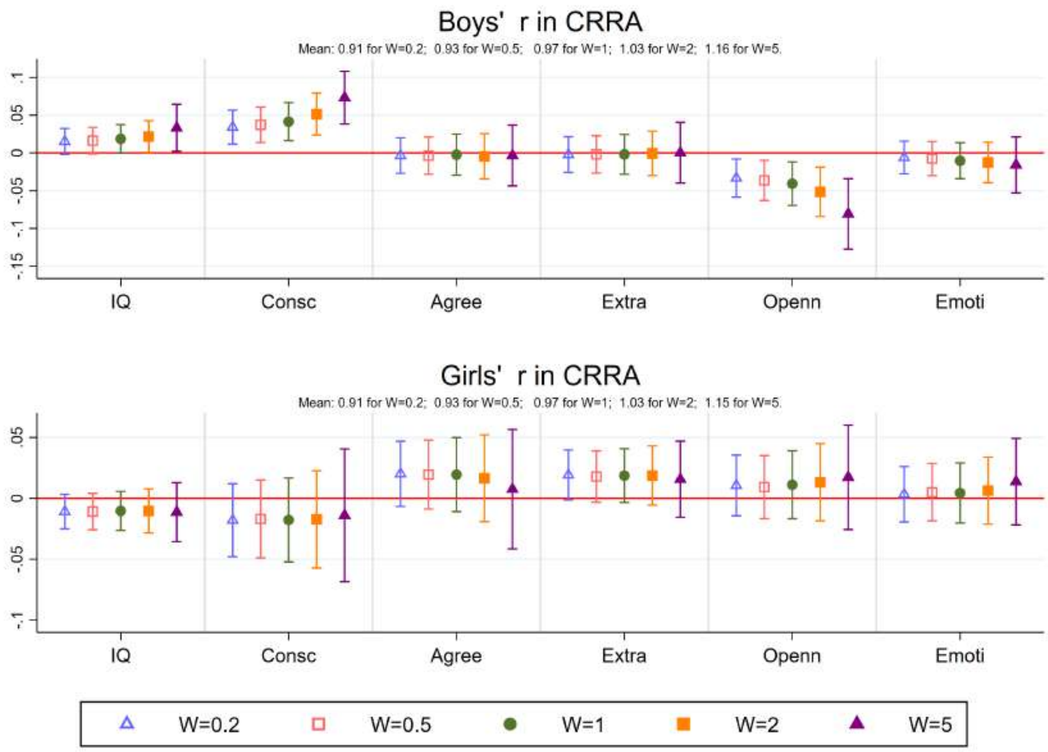


Figure G.5: Risk Aversion r Across Reference Wealth Levels (CRRA)

Notes: We compute predictions for each observation and average them over the sample. Markers denote reference wealth levels: $w = 0.2$ (blue triangle), 0.5 (red square), 1 (green circle, baseline), 2 (orange square), 5 (purple triangle). Mean predictions at each w appear in the panel subtitles. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

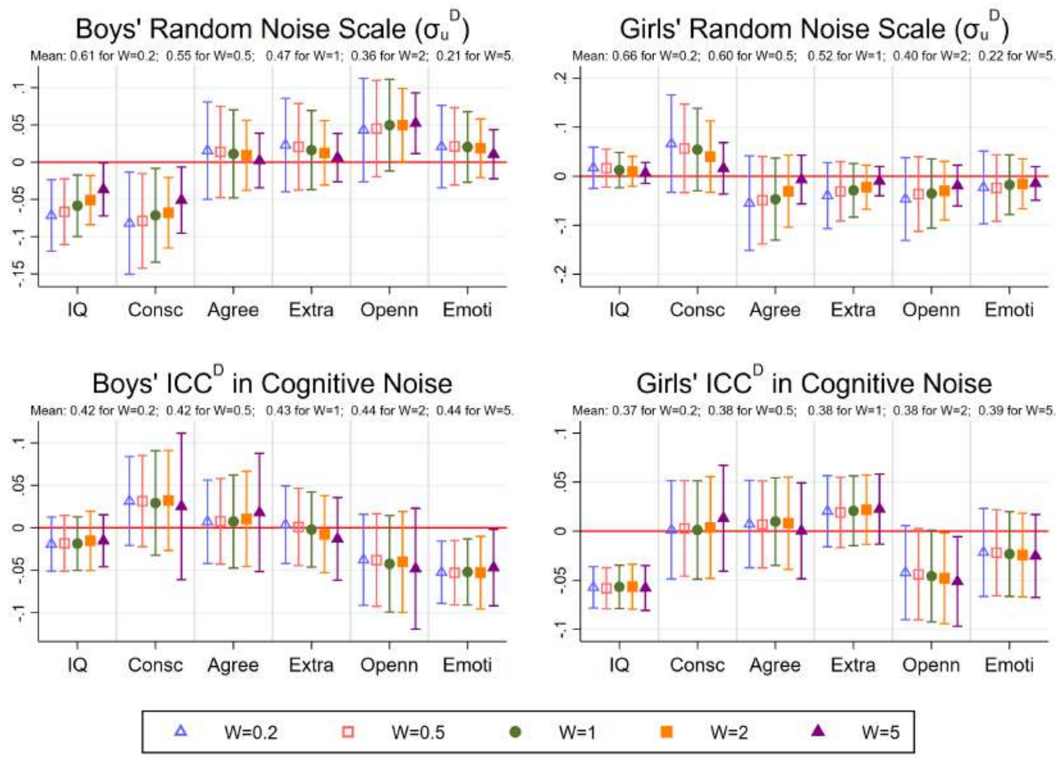


Figure G.6: Transitory Noise σ_{u^D} and ICC^D Across Reference Wealth Levels (CRRA)

Notes: We compute predictions for each observation and average them over the sample. Markers denote reference wealth levels: $w = 0.2$ (blue triangle), 0.5 (red square), 1 (green circle, baseline), 2 (orange square), 5 (purple triangle). Mean predictions at each w appear in the panel subtitles. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

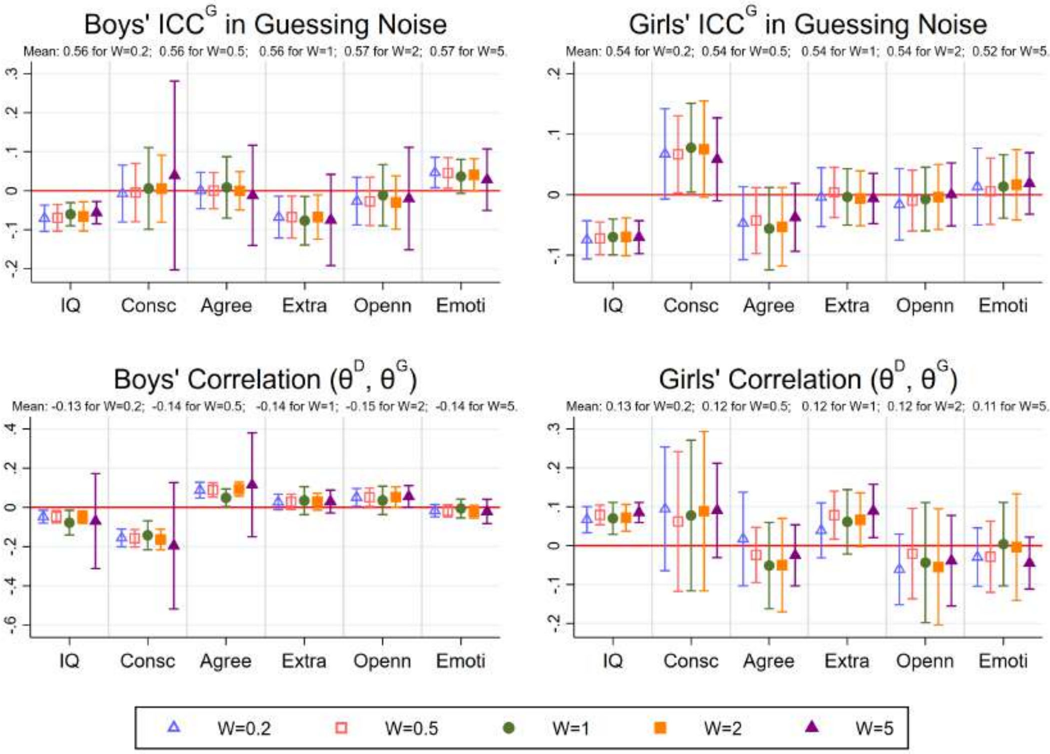


Figure G.7: Guessing Persistence ICC^G and Deliberation–Guessing Correlation Across Reference Wealth Levels (CRRA)

Notes: We compute predictions for each observation and average them over the sample. Markers denote reference wealth levels: $w = 0.2$ (blue triangle), 0.5 (red square), 1 (green circle, baseline), 2 (orange square), 5 (purple triangle). Mean predictions at each w appear in the panel subtitles. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

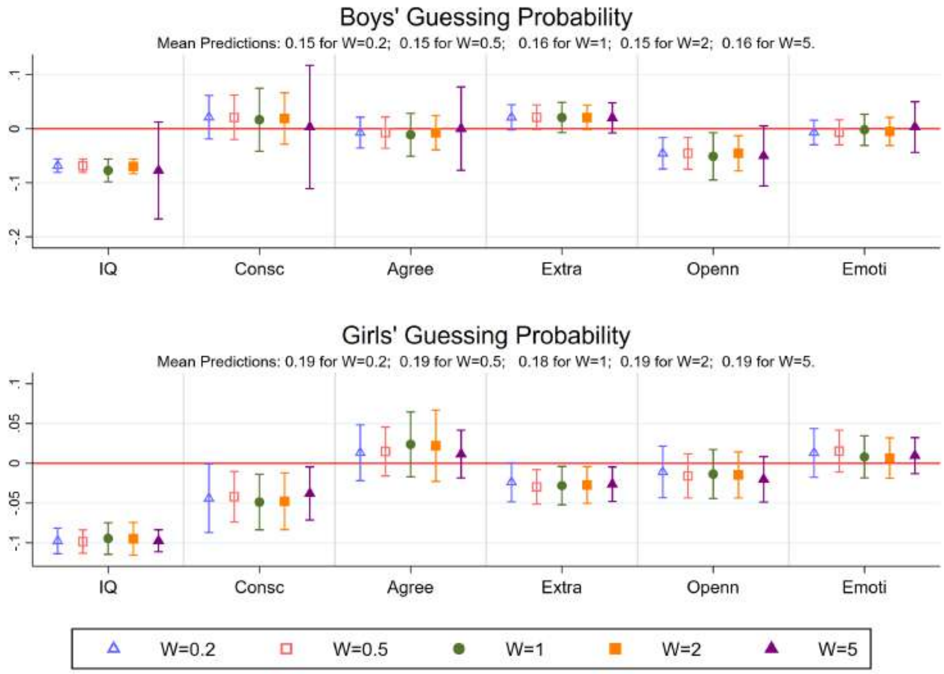


Figure G.8: Guessing Probability Across Reference Wealth Levels (CRRA)

Notes: We compute predictions for each observation and average them over the sample. Markers denote reference wealth levels: $w = 0.2$ (blue triangle), 0.5 (red square), 1 (green circle, baseline), 2 (orange square), 5 (purple triangle). Mean predictions at each w appear in the panel subtitles. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

H Reference Wealth and Socioeconomic Background

This appendix examines whether socioeconomic status affects the reference wealth level in children's utility functions. Children from different socioeconomic backgrounds may develop different reference points based on their lived experiences, parental resources, and consumption exposure. Understanding this heterogeneity is important for interpreting risk preference estimates and their distributional implications.

We compare two model specifications across both Expo-Power and CRRA utility functions. Model 1 fixes reference wealth w at unity for all children, imposing a common reference point. Model 2 estimates w as a flexible function of SES variables including urban residence, parental education (high school completion for both father and mother), sibling presence, and left-behind status (father only or both parents) and the left-behind volatility index.

For each utility specification, we present coefficient plots showing how IQ and personality traits (the Big Five: Conscientiousness, Agreeableness, Extraversion, Openness, and Emotional Stability) affect the estimated parameters under both models. Blue squares represent Model 1 (fixed $w = 1$), red circles represent Model 2 (estimated w from SES). The figures allow comparison of parameter stability across specifications and assessment of whether incorporating SES-dependent reference wealth meaningfully alters the estimated relationships between skills and preference parameters.

Figures H.1–H.5 present results for the Expo-Power specification. Figures H.6–H.10 present corresponding results for the CRRA specification.

Despite allowing reference wealth to vary with background, the coefficient patterns for IQ and personality traits remain qualitatively similar across models. Risk aversion predictions shift modestly—from 0.77 to 0.91 for boys and from 0.58 to 0.52 for girls—while wealth-dependence β is essentially unchanged (about 0.09 for both genders); the relative magnitudes and the pattern of trait effects are preserved. Conscientiousness remains the strongest predictor for boys in both models, while Extraversion maintains its positive effect for girls.

Decision noise parameters show even greater stability. ICC^D is 0.43 for boys in both models and 0.37–0.38 for girls. ICC^G remains at 0.57–0.58 for boys and 0.53–0.54 for girls. The correlation $\rho(\theta^D, \theta^G)$ is -0.15 to -0.16 for boys and 0.11 for girls in both specifications. Guessing probability is 0.14–0.15 for boys and 0.18–0.19 for girls regardless of whether reference wealth is fixed or estimated.

H.1 Expo-Power Specification

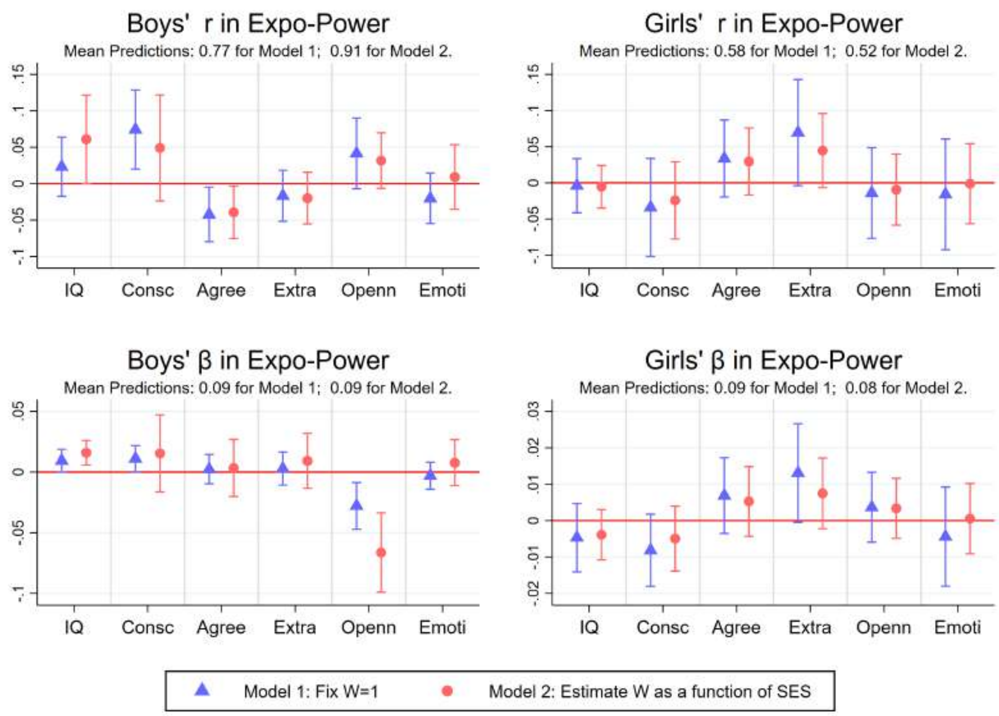


Figure H.1: Reference Wealth: Risk Aversion Parameters r and β (Expo-Power)

Notes: We compute predictions for each observation and average them over the sample. Blue squares: Model 1 (fixed reference wealth $w = 1$). Red circles: Model 2 (w estimated as a function of SES variables). Mean predictions appear in each panel's subtitle. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

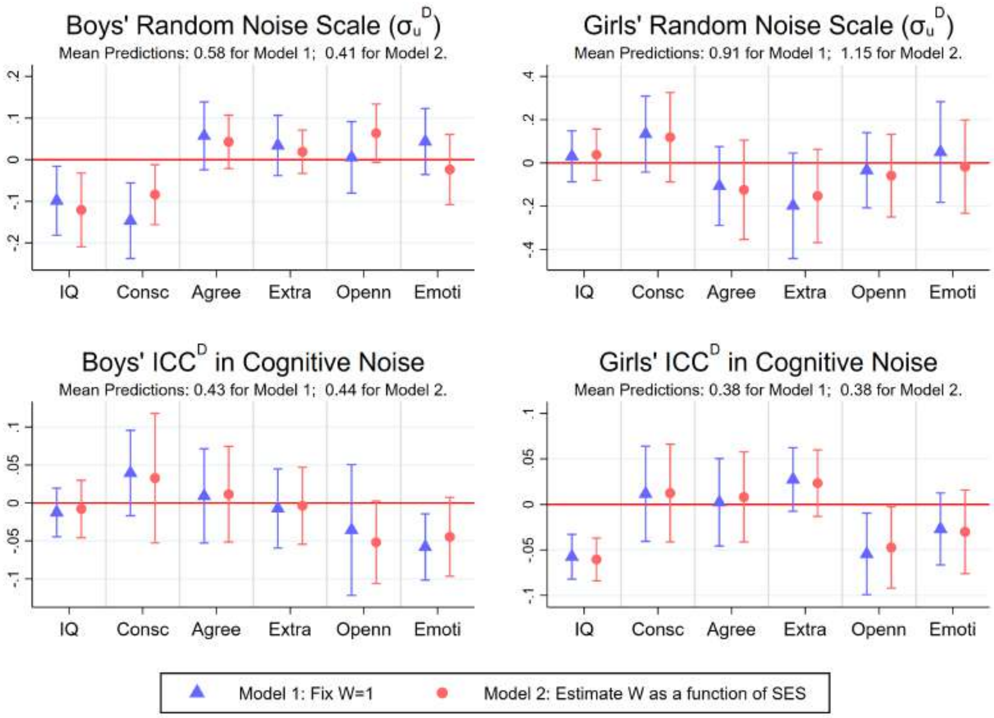


Figure H.2: Reference Wealth: Transitory Noise σ_{u^D} and ICC^D (Expo-Power)

Notes: We compute predictions for each observation and average them over the sample. Blue squares: Model 1 (fixed reference wealth $w = 1$). Red circles: Model 2 (w estimated as a function of SES variables). Mean predictions appear in each panel's subtitle. Error bars show 90% confidence intervals. N = 1,115 boys, N = 1,089 girls.

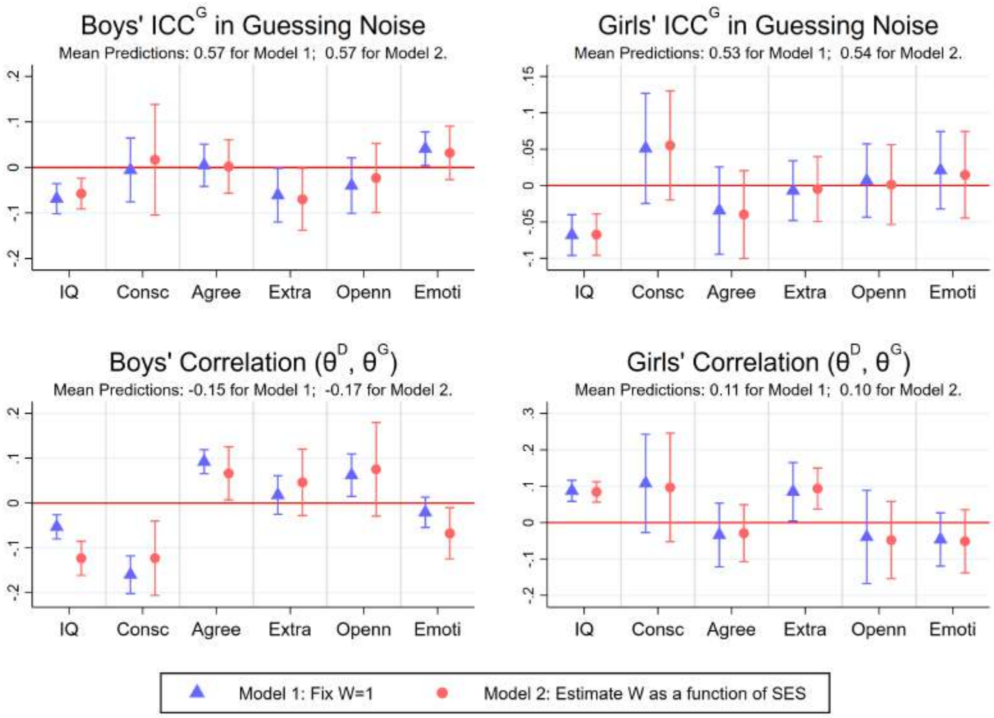


Figure H.3: Reference Wealth: ICC^G and Deliberation–Guessing Correlation (Expo-Power)

Notes: We compute predictions for each observation and average them over the sample. Blue squares: Model 1 (fixed reference wealth $w = 1$). Red circles: Model 2 (w estimated as a function of SES variables). Mean predictions appear in each panel's subtitle. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

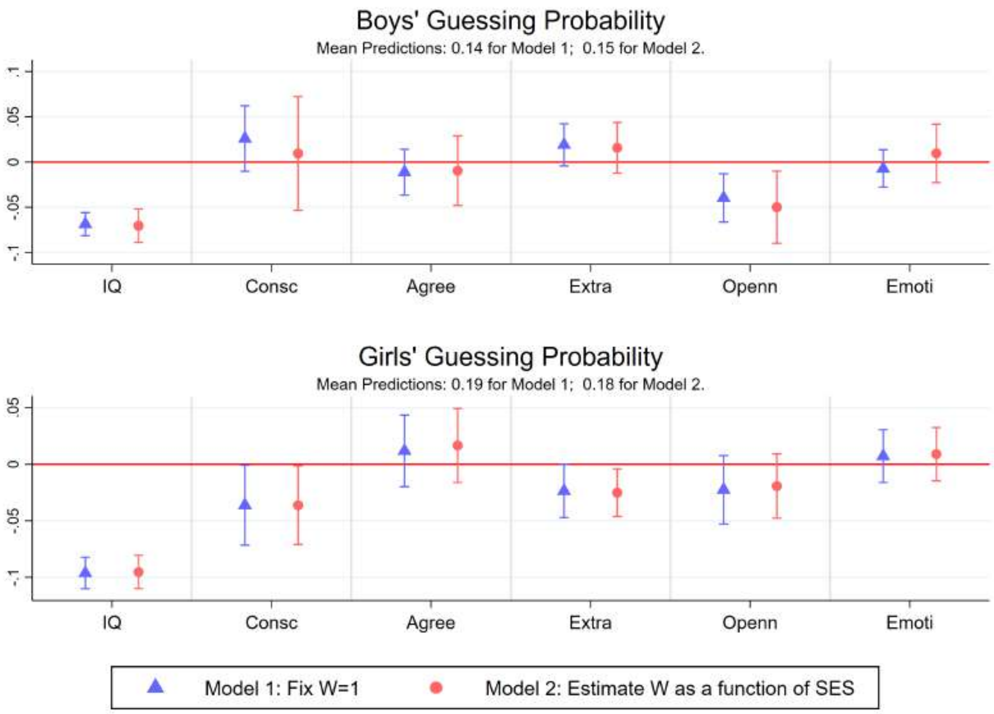


Figure H.4: Reference Wealth: Guessing Probability (Expo-Power)

Notes: We compute predictions for each observation and average them over the sample. Blue squares: Model 1 (fixed reference wealth $w = 1$). Red circles: Model 2 (w estimated as a function of SES variables). Mean predictions appear in each panel's subtitle. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

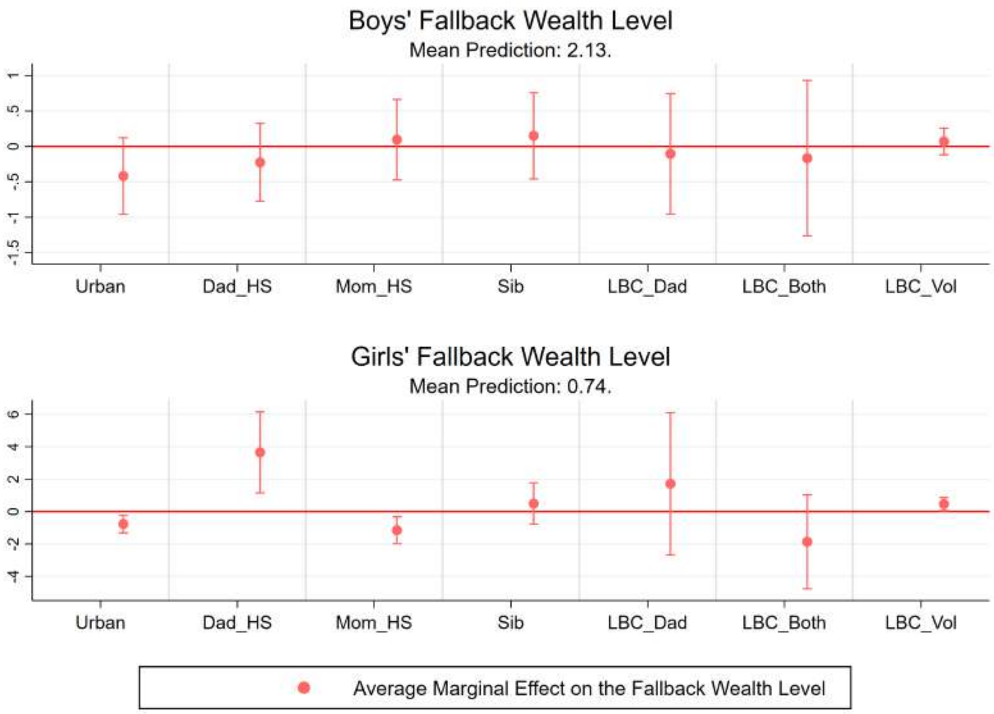


Figure H.5: Reference Wealth: Estimated w Across SES Variables (Expo-Power)

Notes: We compute predictions for each observation and average them over the sample. Averaged over the sampling distribution of the SES variables, the estimated reference wealth is $w = 2.13$ for boys and $w = 0.74$ for girls. The reference wealth level is shown only for Model 2 (w estimated from SES). Model 1 fixes $w = 1$ and is therefore not displayed. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

H.2 CRRA Specification

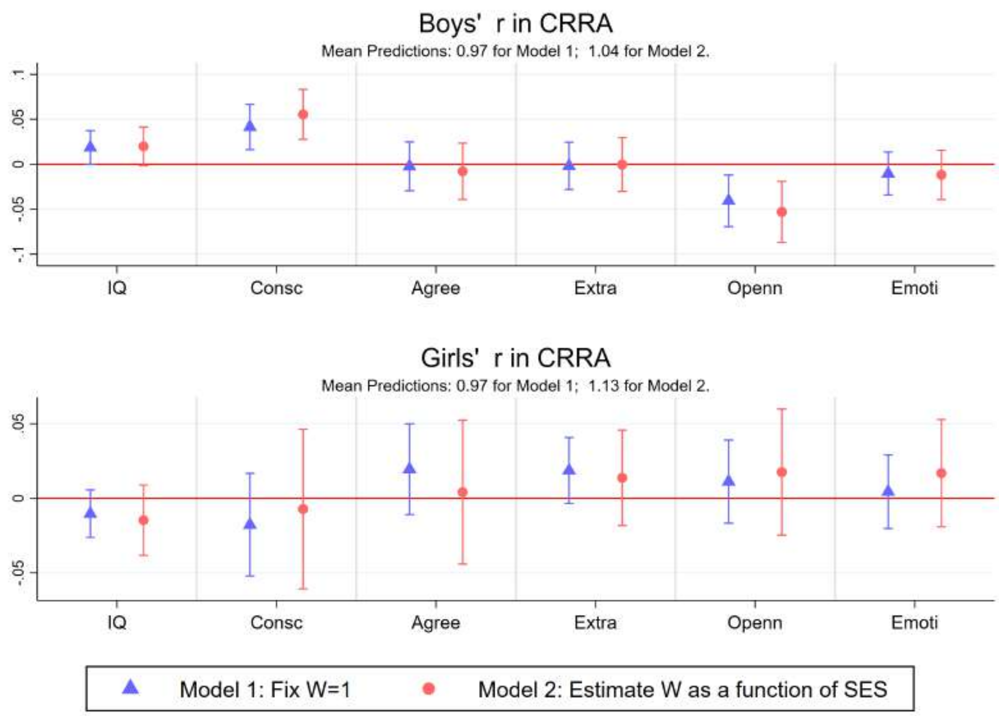


Figure H.6: Reference Wealth: Risk Aversion Parameter r (CRRA)

Notes: We compute predictions for each observation and average them over the sample. Blue squares: Model 1 (fixed reference wealth $w = 1$). Red circles: Model 2 (w estimated as a function of SES variables). Mean predictions appear in each panel's subtitle. Error bars show 90% confidence intervals. N = 1,115 boys, N = 1,089 girls.

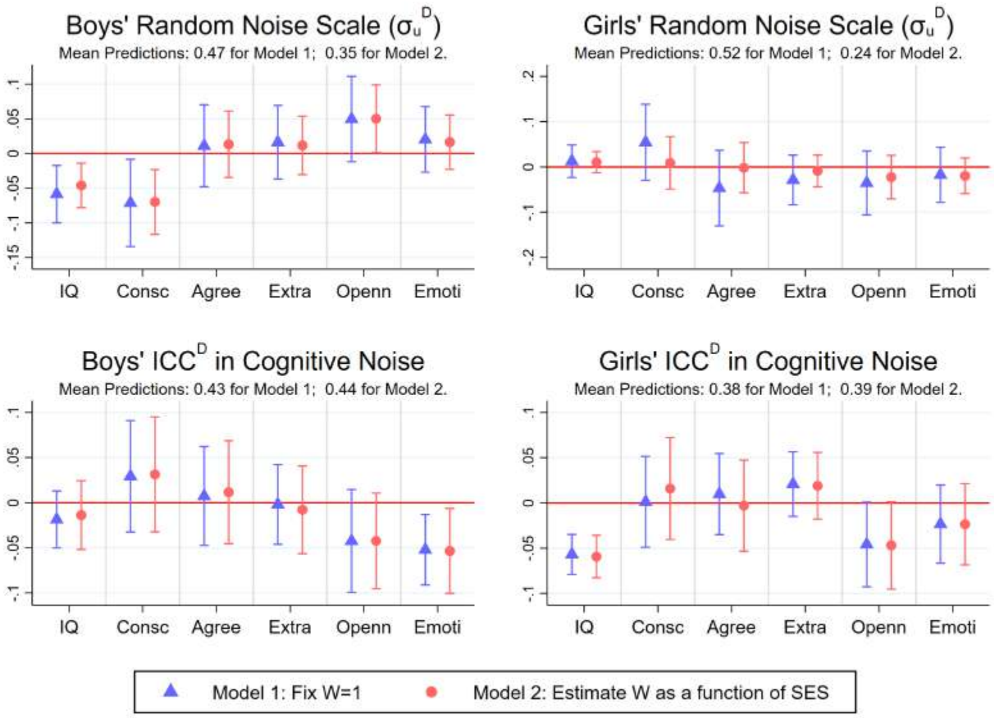


Figure H.7: Reference Wealth: Transitory Noise σ_{u^D} and ICC^D (CRRA)

Notes: We compute predictions for each observation and average them over the sample. Blue squares: Model 1 (fixed reference wealth $w = 1$). Red circles: Model 2 (w estimated as a function of SES variables). Mean predictions appear in each panel's subtitle. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

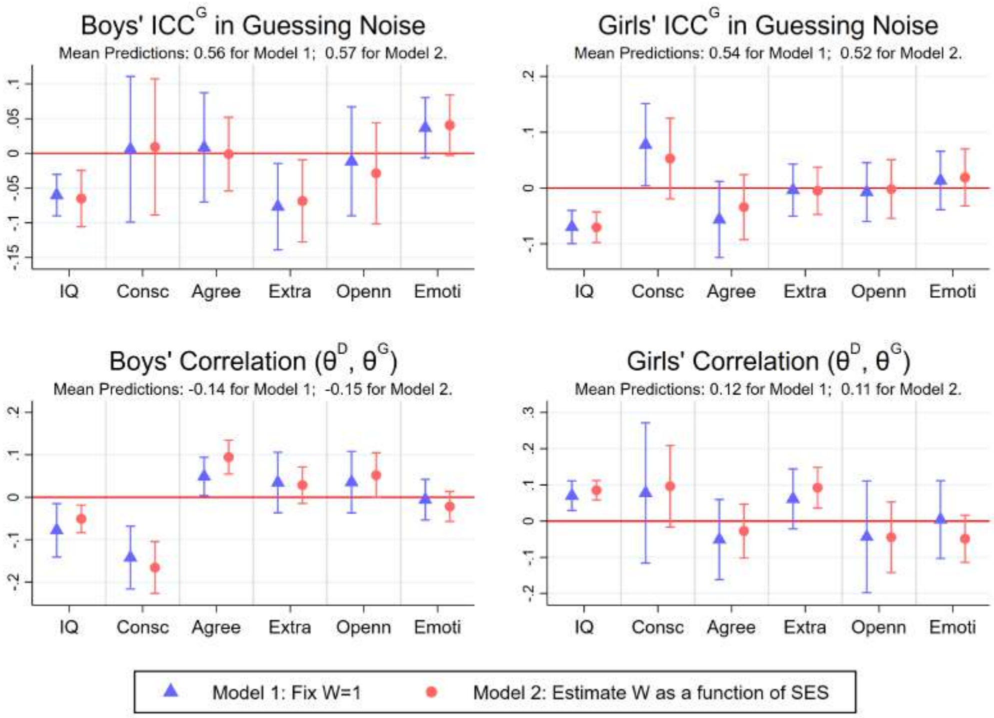


Figure H.8: Reference Wealth: ICC^G and Deliberation–Guessing Correlation (CRRA)

Notes: We compute predictions for each observation and average them over the sample. Blue squares: Model 1 (fixed reference wealth $w = 1$). Red circles: Model 2 (w estimated as a function of SES variables). Mean predictions appear in each panel's subtitle. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

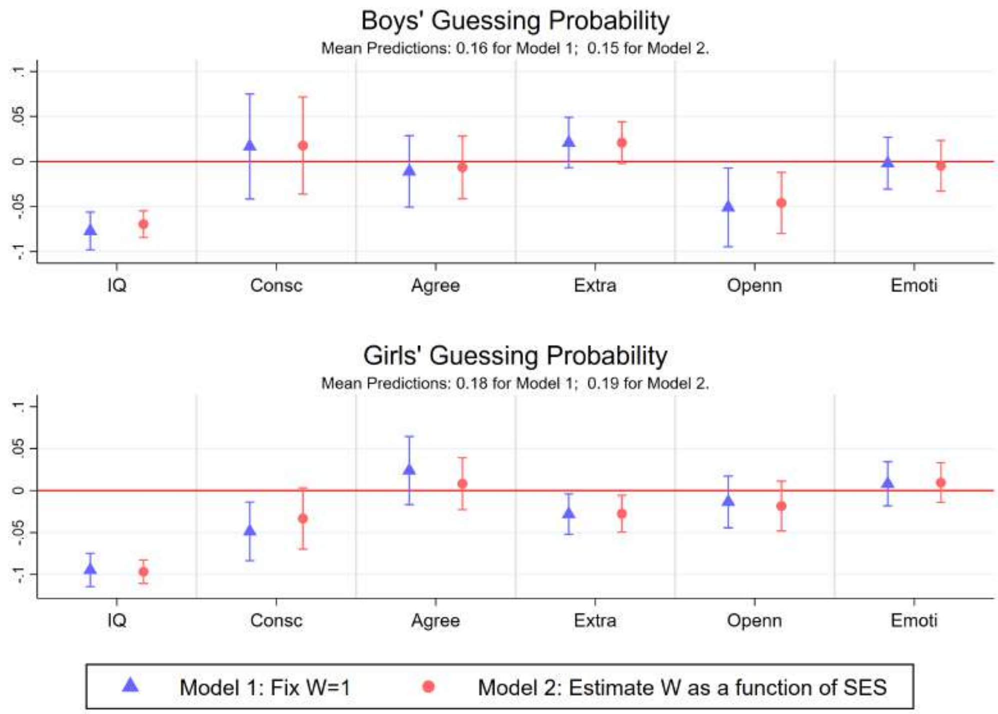


Figure H.9: Reference Wealth: Guessing Probability (CRRA)

Notes: We compute predictions for each observation and average them over the sample. Blue squares: Model 1 (fixed reference wealth $w = 1$). Red circles: Model 2 (w estimated as a function of SES variables). Mean predictions appear in each panel's subtitle. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

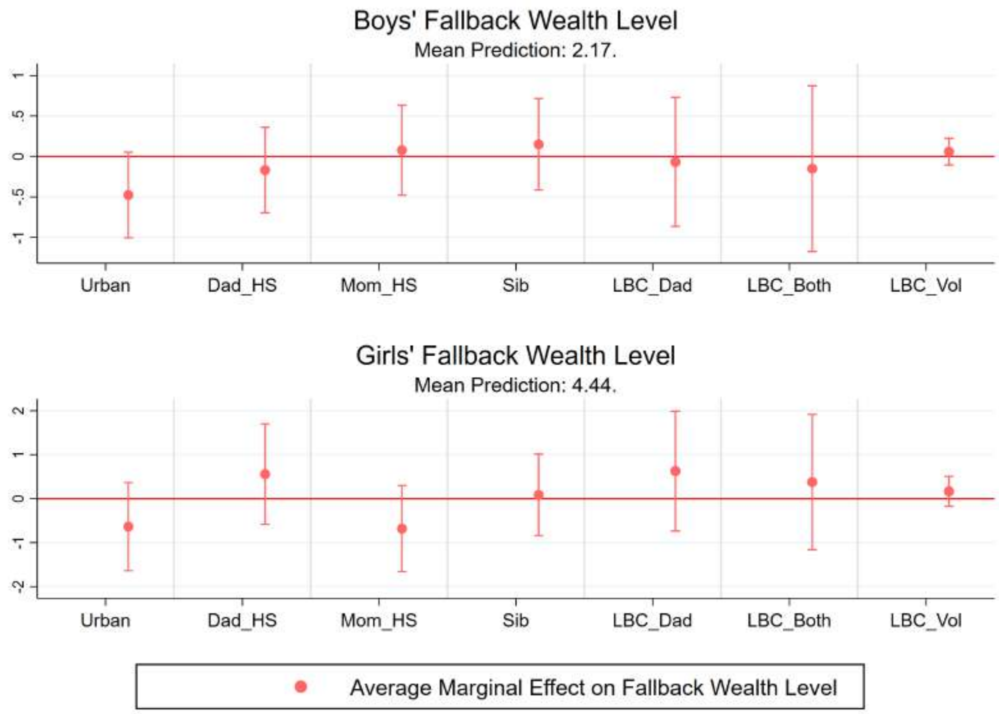


Figure H.10: Reference Wealth: Estimated w Across SES Variables (CRRA)

Notes: We compute predictions for each observation and average them over the sample. The reference wealth level is shown only for Model 2 (w estimated from SES). Model 1 fixes $w = 1$ and is therefore not displayed. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

I Comparing Expo-Power Utility Functions: Xie (2000) and Saha (1993)

This appendix[28] presents a comprehensive comparison between the Expo-Power utility function introduced by Xie (2000) and a related specification by Saha (1993). We characterize conditions under which the utility function exhibits strict concavity in the choice space, which ensures a well-defined interior solution for asset allocation.

[28]We thank Ariyan Mishra for checking the concavity conditions, and we draw on his work. We build on it by relating two alternative specifications of the utility function.

I.1 The Xie (2000) Expo-Power Specification

As presented in Section 4, the Expo-Power utility function introduced by Xie (2000) is:

$$U_{\text{ExpoPower}}(s+w) = \frac{1}{\beta}\left[1 - \exp\left\{-\beta\frac{(s+w)^{1-r}-1}{1-r}\right\}\right], \quad \beta \in \mathbb{R},\ r \geq 0, \tag{I.1}$$

where $s \geq 0$ represents the payoff from the experimental task and $w > 0$ represents reference wealth (consumption level before participation). This specification produces wealth-dependent measures of risk aversion:

$$\text{RRA}(s+w) = r + \beta(s+w)^{1-r}, \tag{I.2}$$

$$\text{ARA}(s+w) = \frac{\text{RRA}(s+w)}{s+w} = \frac{r}{s+w} + \beta(s+w)^{-r}. \tag{I.3}$$

At reference wealth ($w = 1$), the relative risk aversion equals $\text{RRA}(w) = r + \beta w^{1-r} = r + \beta$, which is the parameterization we use throughout the main text. We impose no sign restriction on β: concavity over the relevant range requires only $\text{RRA}(s+w) = r+\beta(s+w)^{1-r} > 0$ (equivalently $\beta(s+w)^{1-r} > -r$), so that decreasing relative risk aversion ($\beta < 0$) remains admissible and is ruled out empirically rather than by assumption.

I.2 Relationship to Saha (1993)

Saha (1993) studied a more restrictive form:

$$U_{\text{Saha}}(z; \alpha, \beta) = \theta - \exp\{-\beta z^{\alpha}\}, \tag{I.4}$$

where $z = s + w$ denotes total wealth and $\alpha = 1 - r$ in Xie's notation. This is the exponential kernel of the Xie specification, but the two exponents are not identical: Xie's is $-\beta\,[(s+w)^{1-r} - 1]/(1-r)$ whereas Saha's is $-\beta(s+w)^{1-r}$. The two therefore coincide only up to the normalizing factor $1/(1-r)$ and the additive constant inside the exponent, so that Saha's β corresponds to Xie's $\beta/(1-r)$. The key difference is that Xie (2000) includes the outer transformation $\frac{1}{\beta}[1 - \exp\{\cdot\}]$, which:

1. Bounds utility from above when $\beta > 0$, introducing satiation
2. Ensures proper scaling for comparison across different β values
3. Provides the correct limiting behavior as $\beta \to 0$ (converging to CRRA)

Throughout this paper, we use the Xie (2000) specification, as it captures the limiting case of the CRRA functional form and facilitates smoother estimation.

J Contributions of the Components to the Full Model

Table 2 in the main text documents how model fit and parameter estimates change as components of the model are added. Panel A shows that each component contributes significantly to model fit, with ICC structure providing the largest improvements (approximately 1,955–1,965 log-likelihood points for boys and 1,768–1,782 for girls; these are single log-likelihood gains, half the corresponding likelihood-ratio statistics). Panel B confirms that Expo-Power dominates CRRA by an AIC criterion in the full specification.

J.1 Components of Deliberation Noise: ICC Structure and Guessing Behavior

A fundamental question is whether decision errors reflect purely independent shocks or contain persistent individual components. We address this by first estimating a model without covariates, incorporating two mechanisms: ICC structure that captures stable individual differences in deliberation noise, and explicit guessing behavior. This baseline reveals the model's core properties before we assess the role of psychological traits and background.

We separately test the null hypotheses of an independent error structure and the absence of guessing behavior using likelihood ratio tests.[29]

Table J.1 presents specifications incorporating each component separately. Panel A examines an ICC structure without guessing. ICC dramatically improves model fit, with likelihood ratio tests decisively rejecting independence (boys: LR > 3,900, girls: LR > 3,500, both $p < 0.001$). The ICC^D parameters—0.178–0.180 for boys and 0.161–0.163 for girls—indicate that approximately 16–18% of deliberation noise variance stems from stable individual traits rather than task-specific shocks in this baseline specification (this share rises to approximately 40% in the full model with guessing and traits). Critically, Expo-Power maintains its advantage over CRRA even after both include ICC (boys: LR = 16.2, $p < 0.001$, girls: LR = 182.3, $p < 0.001$).

The deliberation noise parameter (σ_{u^D}) responds differently to ICC inclusion across specifications. For CRRA, adding ICC reduces noise substantially: from 1.368 to 0.895 for boys (35% reduction) and from 1.056 to 0.965 for girls (9% reduction), confirming that baseline CRRA models misattribute persistent heterogeneity to transitory noise. For Expo-Power with ICC, noise estimates are 1.046 for boys and 1.720 for girls. The noise reduction is much larger in our full model, which incorporates ICC, guessing, the effects of IQ and personality traits, and the flexible Expo-Power utility specification.

Appendix L demonstrates that, in the full-model setting, omitting ICC^D inflates the transitory noise (Appendix figure L.2), underestimates risk aversion—through wealth-dependence β under Expo-Power and through curvature r under CRRA (Appendix figures L.1 and L.5)—and misallocates guessing to transitory noise (Appendix figure L.4), confirming that a substantial share of apparent noise reflects stable individual differences. Appendix M further shows that omitting ICC attenuates the effects of traits on preferences. Without ICC^D, the Conscientiousness effect on risk aversion is substantially attenuated

[29] Due to boundary constraints in the parameter space for ICC and guessing, the log-likelihood ratios associated with these null hypotheses follow a mixed χ^2 distribution: a point mass at 0 with probability 1/2 and a χ^2 distribution with 1 degree of freedom with probability 1/2. See Chernoff (1954) and Self and Liang (1987). We apply the same approximation in specifications that control for covariates.

and loses significance, demonstrating that proper error structure is essential for recovering trait-preference relationships (Appendix figure M.1).

Panel B of Table J.1 examines guessing behavior without ICC^D. Beyond deliberation noise, children may sometimes guess, responding randomly. Adding guessing improves model fit (boys: LR = 50.6–57.5, girls: LR = 3.6–27.5), but substantially less than adding ICC structure. Nevertheless, the likelihood ratio tests reject the null hypothesis of no guessing behavior at the 5% level (p-value = 0.000 for boys, 0.029 for girls). The estimated mean guessing probability varies by specification and gender: under CRRA, boys show guessing rates of 15.6% while girls show 6.7%. Under Expo-Power, boys show 16.4% and girls show 14.3%.

Expo-Power maintains superiority over CRRA after incorporating guessing (boys: LR = 8.9, $p = 0.003$, girls: LR = 178.5, $p < 0.001$). Appendix N demonstrates that in the full model, guessing captures qualitatively distinct decision processes: models omitting guessing inflate the transitory noise by over 110% (Appendix figure N.2) and collapse ICC^D by roughly 55%, demonstrating that guessing operates through both persistent and transitory channels.

Table J.1: Model Estimates: Correlated Shocks and Guessing

	Boys			Girls		
	CRRA No ICC	CRRA +ICC	Expo-Power +ICC	CRRA No ICC	CRRA +ICC	Expo-Power +ICC
Panel A: No Guessing, Without IQ or Traits						
r in Expo-Power + CRRA	0.845***	0.939***	0.760***	0.946***	0.949***	0.553***
	(0.014)	(0.010)	(0.034)	(0.013)	(0.009)	(0.024)
β in Expo-Power	—	—	0.082***	—	—	0.086***
			(0.007)			(0.004)
σ_{u^D}	1.368***	0.895***	1.046***	1.056***	0.965***	1.720***
	(0.073)	(0.038)	(0.072)	(0.053)	(0.037)	(0.134)
ICC^D	—	0.178***	0.180***	—	0.161***	0.163***
		(0.010)	(0.010)		(0.009)	(0.009)
Loglikelihood	−36015.7	−34059.2	−34051.1	−36012.6	−34244.7	−34153.6
LR: $+ICC^D$		3913.0***	3927.2***		3535.8***	3563.4***
p-value		[0.000]	[0.000]		[0.000]	[0.000]
LR: Expo-Power vs. CRRA			16.2***			182.3***
p-value			[0.000]			[0.000]
Panel B: No ICC, Without IQ or Traits						
	CRRA No Guess	CRRA +Guess	Expo-Power +Guess	CRRA No Guess	CRRA +Guess	Expo-Power +Guess
r in Expo-Power + CRRA	0.845***	0.843***	0.711***	0.946***	0.948***	0.527***
	(0.014)	(0.015)	(0.041)	(0.013)	(0.014)	(0.024)
β in Expo-Power	—	—	0.047***	—	—	0.078***
			(0.009)			(0.004)
σ_{u^D}	1.368***	1.062***	1.227***	1.056***	0.956***	1.678***
	(0.073)	(0.066)	(0.098)	(0.053)	(0.075)	(0.141)
Mean Prob(Guess)	—	0.156***	0.164***	—	0.067**	0.143***
		(0.018)	(0.018)		(0.033)	(0.024)
Loglikelihood	−36015.7	−35990.4	−35986.0	−36012.6	−36010.8	−35921.6
LR: +Guess		50.6***	57.5***		3.6**	27.5***
p-value		[0.000]	[0.000]		[0.029]	[0.000]
LR: Expo-Power vs. CRRA			8.9***			178.5***
p-value			[0.003]			[0.000]

Notes: Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Panel A shows ICC effects without guessing. Panel B shows guessing effects without ICC. All models in this table exclude IQ, personality traits, and background variables. ICC$^D = \sigma^2_{\theta^D}/(\sigma^2_{u^D} + \sigma^2_{\theta^D})$ measures the proportion of variance in deliberation noise attributable to its persistent component. Because ICC and guessing are tested at the boundary of the parameter space, the asymptotic p-values of the "LR: $+ICC^D$" and "LR: +Guess" tests are based on a mixed distribution with a point mass at 0 with probability 1/2 and a χ^2 distribution with 1 degree of freedom with probability 1/2 (Chernoff 1954; Self and Liang 1987). Within each panel, the "LR: $+ICC^D$" and "LR: +Guess" statistics compare each model against the corresponding baseline of the *same* utility specification (CRRA against CRRA, Expo-Power against Expo-Power). The "LR: Expo-Power vs. CRRA" statistics compare the two utility specifications and are based on a regular χ^2 distribution.

J.2 The Impact of Deliberation Noise and Choice of Functional Form on Parameter Estimates

The first channel we consider operates through deliberation noise. From Table J.2, moving from baseline CRRA to full CRRA (columns 1 vs. 2 for boys, columns 4 vs. 5 for girls) adds ICC structure, guessing behavior, and covariates while maintaining the CRRA functional form. For boys, this change increases r from 0.845 to 0.970—a 15% increase. For girls, the increase is smaller: from 0.946 to 0.968, just 2%. This asymmetric response to deliberation-noise controls narrows the apparent gender gap: under baseline CRRA, girls appear 12% more risk averse than boys, but under the full CRRA model the gap disappears entirely.

The second channel operates through the more flexible functional form. Moving from the fully specified CRRA model to the fully specified Expo-Power model (columns 2 vs. 3 for boys, columns 5 vs. 6 for girls) relaxes the constant-curvature restriction and separates risk aversion curvature r from wealth-dependence β. For boys, risk aversion curvature falls from $r = 0.970$ (full CRRA) to $r = 0.771$ (full Expo-Power); for girls it falls much more, from $r = 0.968$ to $r = 0.576$. The wealth-dependence parameter β, by contrast, is nearly identical across genders (0.091 for boys, 0.093 for girls). The gender differences are therefore driven by risk aversion curvature, not by wealth-dependence: once the CRRA restriction is relaxed, girls' curvature drops far more than boys'. Boys have r roughly one-third higher than girls.

Deliberation noise shows dramatic reductions when correlated shocks are introduced. For boys, σ_{u^D} drops from 1.368 (baseline) to 0.469 (full CRRA) to 0.577 (full Expo-Power). For girls, σ_{u^D} falls from 1.056 to 0.519 to 0.908. The fact that noise estimates differ between CRRA and Expo-Power in the full model indicates that functional form restrictions affect not only preference estimates but also estimates of deliberation noise.

The ICC^D estimates show that approximately 40% of decision variance reflects stable individual traits. For deliberation, ICC^D equals 0.433 (boys) and 0.376 (girls). For guessing, ICC^G equals 0.563 (boys) and 0.540 (girls). The higher ICC in guessing suggests guessing behavior is more trait-like than deliberation noise. Boys guess on approximately 14–16% of trials, and girls guess on 18–19%.

Table J.2: From Baseline to Full Model: CRRA vs. Expo-Power

	Boys			Girls		
	Baseline CRRA	Full CRRA	Full Expo-Power	Baseline CRRA	Full CRRA	Full Expo-Power
r in Expo-Power + CRRA	0.845^{***}	0.970^{***}	0.771^{***}	0.946^{***}	0.968^{***}	0.576^{***}
	(0.014)	(0.012)	(0.033)	(0.013)	(0.011)	(0.025)
β in Expo-Power	—	—	0.091^{***}	—	—	0.093^{***}
			(0.009)			(0.005)
σ_{u^D}	1.368^{***}	0.469^{***}	0.577^{***}	1.056^{***}	0.519^{***}	0.908^{***}
	(0.073)	(0.028)	(0.046)	(0.053)	(0.028)	(0.082)
ICC^D	—	0.433^{***}	0.428^{***}	—	0.376^{***}	0.383^{***}
		(0.025)	(0.033)		(0.019)	(0.019)
ICC^G	—	0.563^{***}	0.574^{***}	—	0.540^{***}	0.531^{***}
		(0.024)	(0.025)		(0.024)	(0.023)
$\rho(\theta^D, \theta^G)$	—	-0.142^{***}	-0.150^{***}	—	0.125^{**}	0.108^{**}
		(0.039)	(0.034)		(0.056)	(0.045)
Mean Prob(Guess)	—	0.158^{***}	0.144^{***}	—	0.184^{***}	0.187^{***}
		(0.019)	(0.014)		(0.015)	(0.012)
Loglikelihood	-36015.707	-32777.287	-32754.608	-36012.605	-33333.503	-33232.961
Δ vs. Baseline	—	3238.4	3261.1	—	2679.1	2779.6
LR: Expo-Power vs. CRRA			45.358^{***}			201.084^{***}
p-value			[0.000]			[0.000]
AIC	72035.414	65638.574	65607.216	72029.210	66751.006	66563.922

Notes: Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. "Baseline CRRA" model excludes IQ and personality traits, ICC structure, and guessing behavior. "Full" model includes guessing behavior, ICC structures for both deliberation and guessing, guessing-deliberation correlation, and IQ and personality traits in each parameter. The r row reports the risk aversion curvature, averaged over the sampling distribution of IQ and personality traits; for CRRA this is the single risk-aversion parameter and $\beta = 0$ by construction, while for Expo-Power r is the risk aversion curvature and β governs wealth-dependence. LR test for Expo-Power vs. CRRA has 7 degrees of freedom (one β parameter plus six trait interactions).

K Baseline CRRA Model

We begin with the simplest specification: CRRA utility without ICC structure, guessing behavior, or psychological traits. This baseline replicates conventional approaches in the literature and establishes a starting point for our generalizations.

Table K.1 presents estimates separately by gender. Girls exhibit significantly higher risk aversion than boys ($r = 0.946$ vs. $r = 0.845$), a 12% difference consistent with the conventional finding across five decades of research. Deliberation noise is lower for girls ($\sigma_{u^D} = 1.056$) than boys ($\sigma_{u^D} = 1.368$), suggesting girls make more precise decisions under this specification. However, as we demonstrate below, these baseline estimates are substantially biased by the failure to account for persistent individual differences in deliberation noise and functional form restrictions.

Table K.1: Baseline CRRA Model

	Boys	Girls
r (risk aversion)	0.845^{***}	0.946^{***}
	(0.014)	(0.013)
σ_{u^D} (deliberation noise)	1.368^{***}	1.056^{***}
	(0.073)	(0.053)
Log-likelihood	−36015.707	−36012.605
N	1,115	1,089

Notes: Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. This baseline specification excludes an ICC structure, guessing behavior, and IQ and personality traits.

L Model Without ICC Structure

This appendix presents results from specifications that include guessing behavior but exclude ICC structure. By comparing these estimates to the full model with ICC (Figures P.1–P.4 in Appendix P), we demonstrate the critical role of ICC in accurately recovering preference parameters and deliberation noise measures. Models without ICC conflate persistent individual traits with transitory noise, leading to severely inflated noise estimates and attenuated trait effects. A substantial share of apparent "noise" in standard models reflects stable individual differences rather than random decision errors, so that ICC structure is essential for preference recovery, not merely a statistical refinement.

L.1 Expo-Power Specification

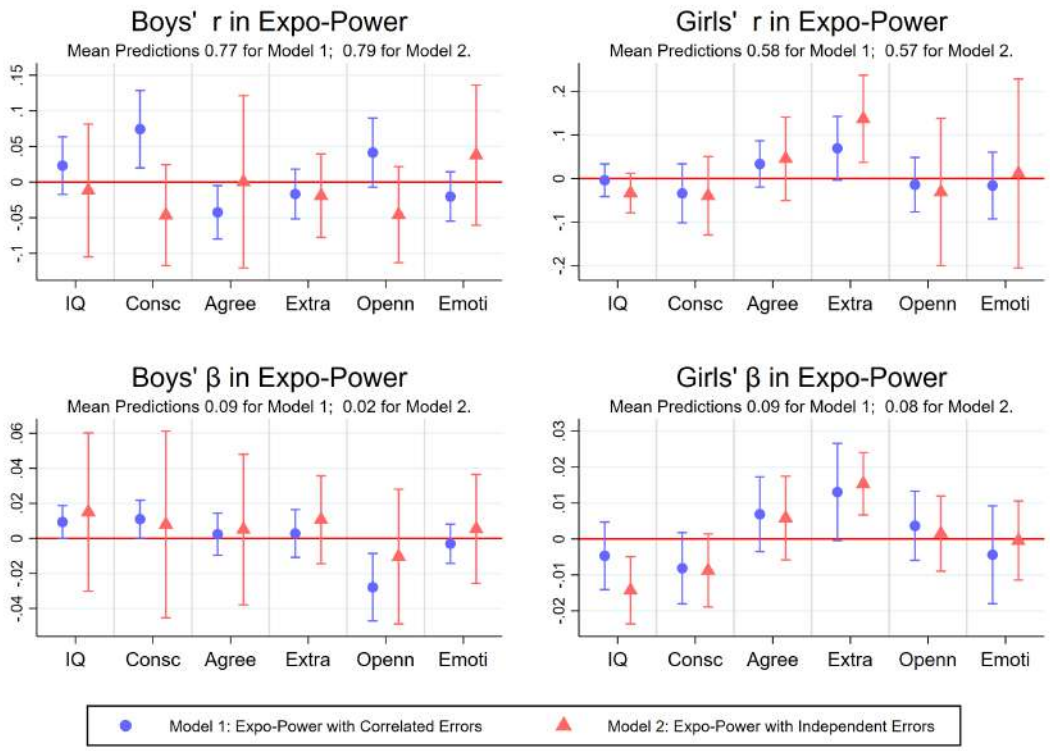


Figure L.1: The Role of ICC: Risk Aversion Parameters r and β (Expo-Power)

We compute predictions for each observation and average them over the sample. Effects of IQ and personality traits on risk aversion curvature r (top panels) and wealth-dependence β (bottom panels). At the mean, omitting ICC leaves r essentially unchanged (boys 0.77 vs. 0.79; girls 0.58 vs. 0.57) but lowers β (boys 0.09 vs. 0.02; girls 0.09 vs. 0.08). Because Expo-Power relative risk aversion $\text{RRA}(s+w) = r + \beta(s+w)^{1-r}$ depends on both parameters, the decline in β underestimates risk aversion. Blue circles: Model 1 with ICC (correlated errors across items). Red triangles: Model 2 without ICC (independent errors). Error bars show 90% confidence intervals. N = 1,115 boys, N = 1,089 girls.

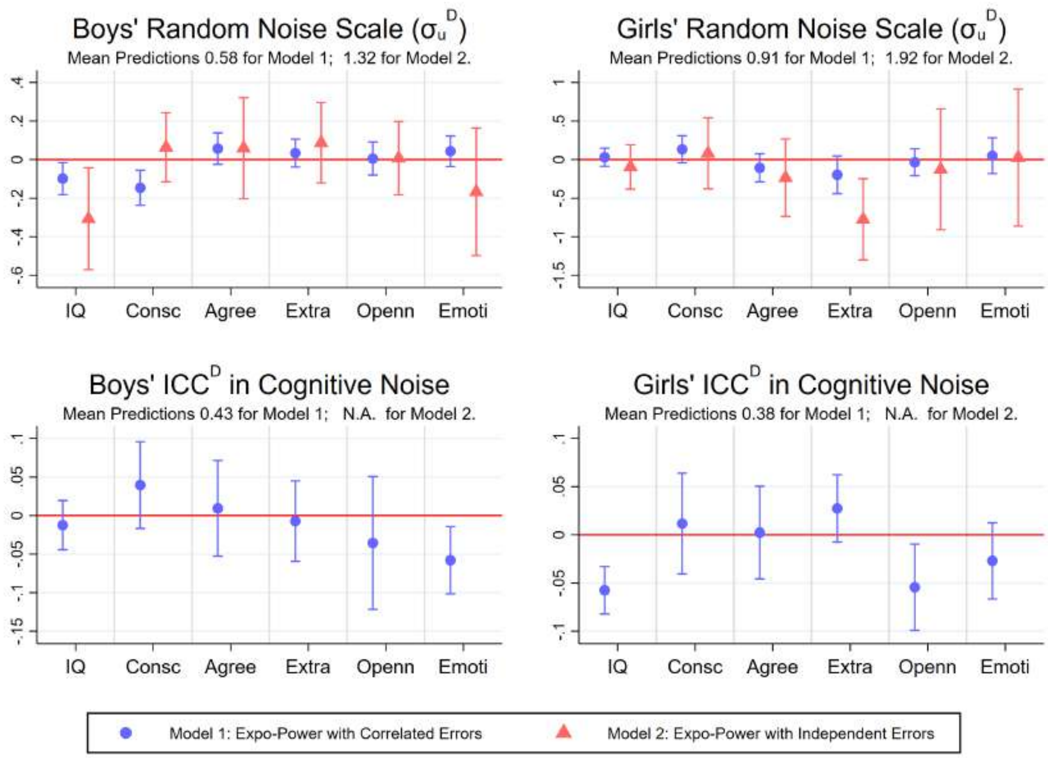


Figure L.2: The Role of ICC: Transitory Noise σ_{u^D} and ICC^D (Expo-Power)

We compute predictions for each observation and average them over the sample. Top panels: transitory noise σ_{u^D}. Bottom panels: ICC^D. At the mean, σ_{u^D} rises when ICC is removed (boys 0.58 vs. 1.32; girls 0.91 vs. 1.92). This rise is partly mechanical: with ICC, part of the deliberation error is held in the persistent component, and removing ICC reallocates it to the transitory term. Comparing total deliberation error—$\sigma^2_{u^D}/(1 - \text{ICC}^D)$ under the model with ICC versus $\sigma^2_{u^D}$ without—the total still increases but by considerably less than σ_{u^D} alone. ICC^D is not estimable without ICC (N.A.). Blue circles: Model 1 with ICC (correlated errors across items). Red triangles: Model 2 without ICC (independent errors). Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

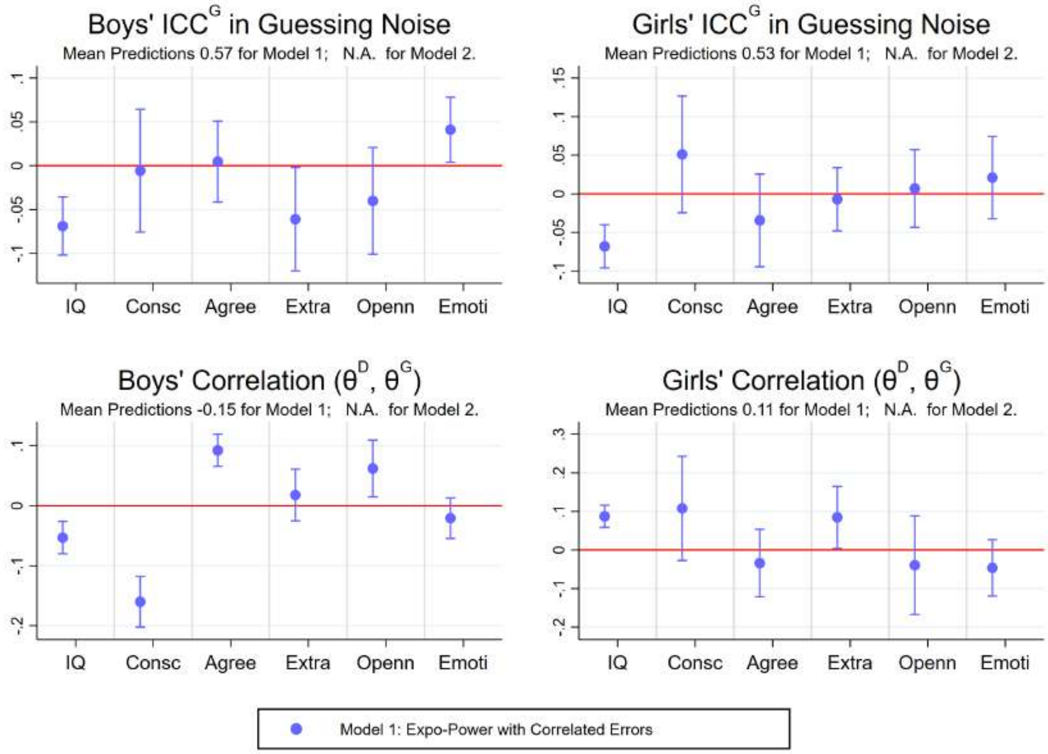


Figure L.3: The Role of ICC: Guessing Persistence ICC^G and Deliberation–Guessing Correlation (Expo-Power)

We compute predictions for each observation and average them over the sample. Top panels: ICC^G. Bottom panels: the correlation $\rho(\theta^D, \theta^G)$ between the persistent deliberation and guessing components. Only Model 1 (with ICC) estimates these parameters, which Model 2 cannot (N.A.). At the mean, ICC$^G = 0.57$ (boys) and 0.53 (girls), and $\rho = -0.15$ (boys) and $+0.11$ (girls). Blue circles: Model 1 with ICC (correlated errors across items). Red triangles: Model 2 without ICC (independent errors). Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

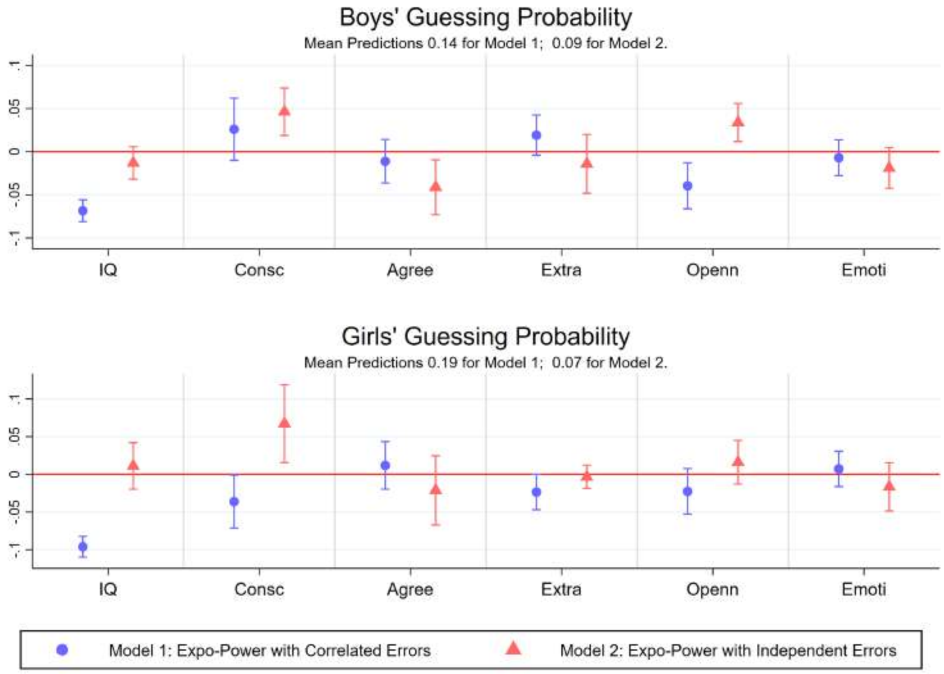


Figure L.4: The Role of ICC: Guessing Probability (Expo-Power)

We compute predictions for each observation and average them over the sample. Effects of IQ and personality traits on the guessing probability. Top panel: boys, bottom panel: girls. At the mean, omitting ICC underestimates guessing (boys 0.14 vs. 0.09; girls 0.19 vs. 0.07), as persistent guessing is reallocated to transitory noise. Blue circles: Model 1 with ICC (correlated errors across items). Red triangles: Model 2 without ICC (independent errors). Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

L.2 CRRA Specification

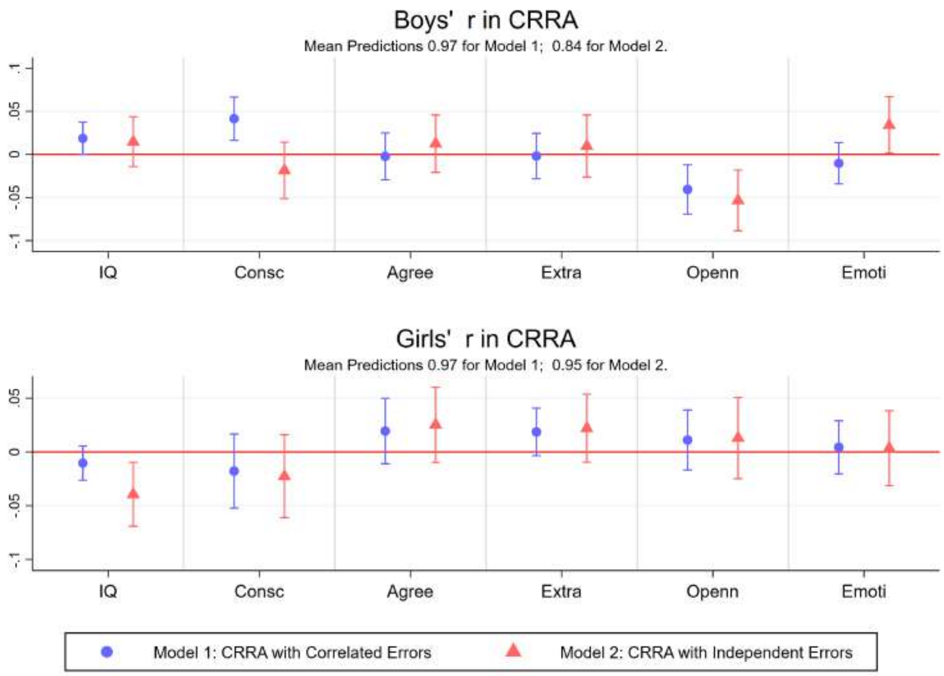


Figure L.5: The Role of ICC: Risk Aversion r (CRRA)

We compute predictions for each observation and average them over the sample. Effects of IQ and personality traits on the CRRA risk aversion curvature r. Top panel: boys, bottom panel: girls. At the mean, omitting ICC lowers r (boys 0.97 vs. 0.84, a 13% underestimate, and girls 0.97 vs. 0.95). Under CRRA risk aversion is underestimated directly through r, the counterpart of the Expo-Power channel that operates through β. Blue circles: Model 1 with ICC (correlated errors across items). Red triangles: Model 2 without ICC (independent errors). Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

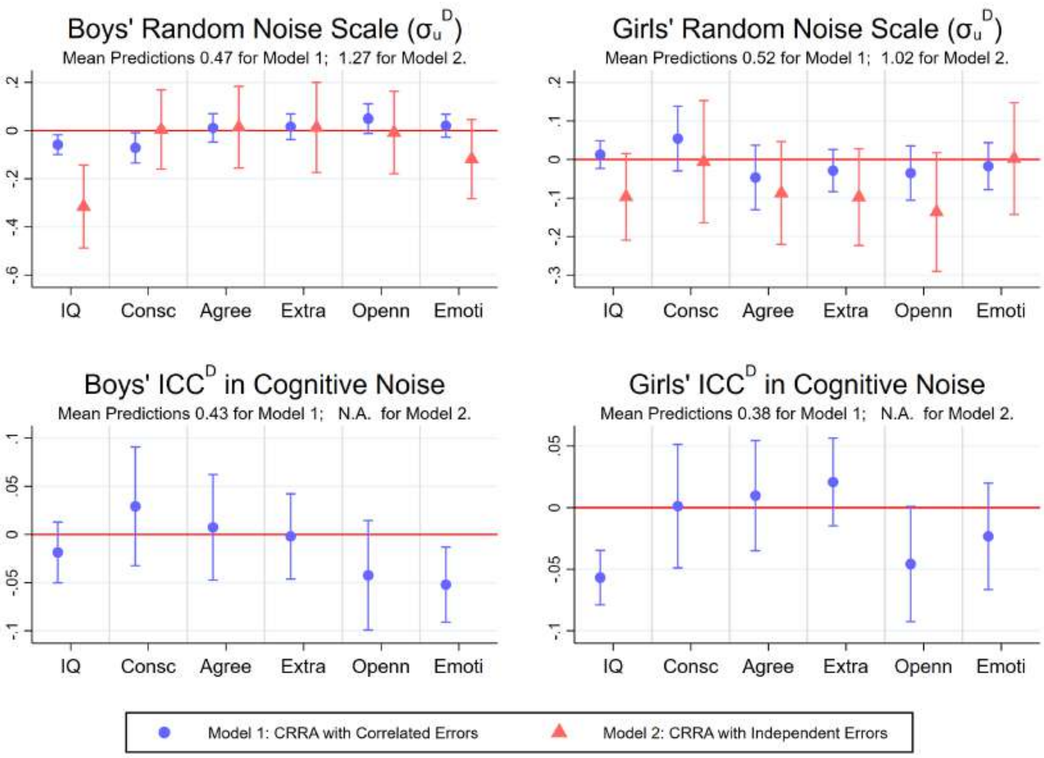


Figure L.6: The Role of ICC: Transitory Noise σ_{u^D} and ICC^D (CRRA)

We compute predictions for each observation and average them over the sample. Top panels: σ_{u^D}. Bottom panels: ICC^D. At the mean, σ_{u^D} rises when ICC is removed (boys 0.47 vs. 1.27; girls 0.52 vs. 1.02), partly mechanically as in the Expo-Power case. ICC^D is not estimable without ICC (N.A.). Blue circles: Model 1 with ICC (correlated errors across items). Red triangles: Model 2 without ICC (independent errors). Error bars show 90% confidence intervals. N = 1,115 boys, N = 1,089 girls.

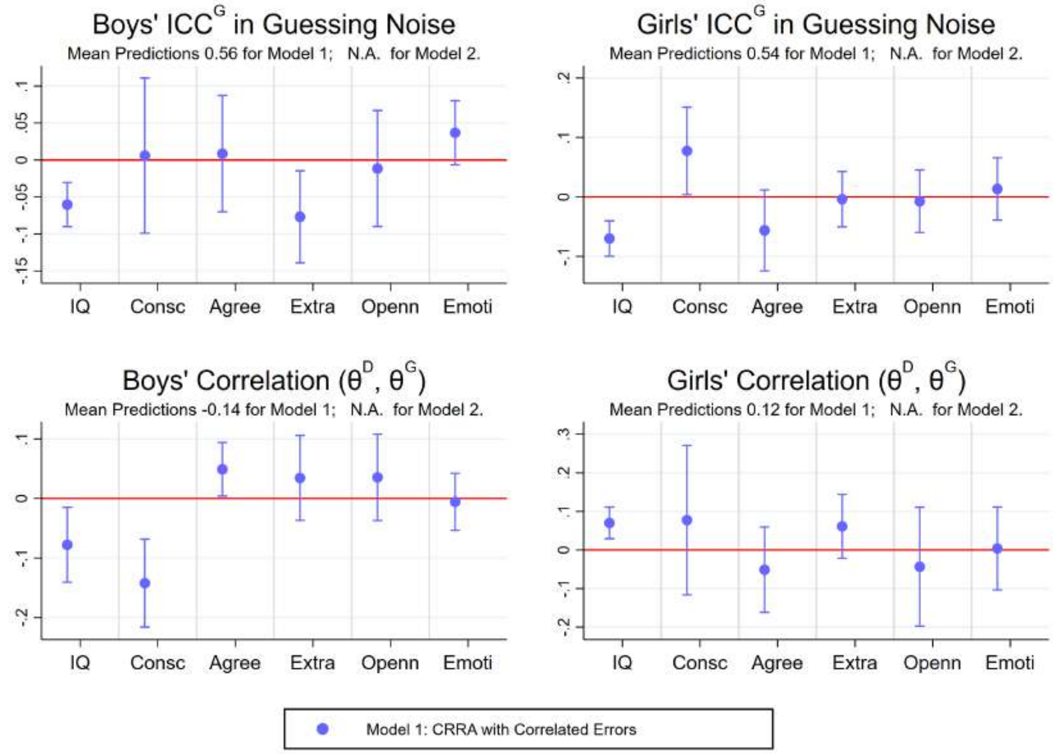


Figure L.7: The Role of ICC: Guessing Persistence ICC^G and Deliberation–Guessing Correlation (CRRA)

We compute predictions for each observation and average them over the sample. Top panels: ICC^G. Bottom panels: $\rho(\theta^D, \theta^G)$. Only Model 1 (with ICC) estimates these parameters. The values closely match the Expo-Power estimates. Blue circles: Model 1 with ICC (correlated errors across items). Red triangles: Model 2 without ICC (independent errors). Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

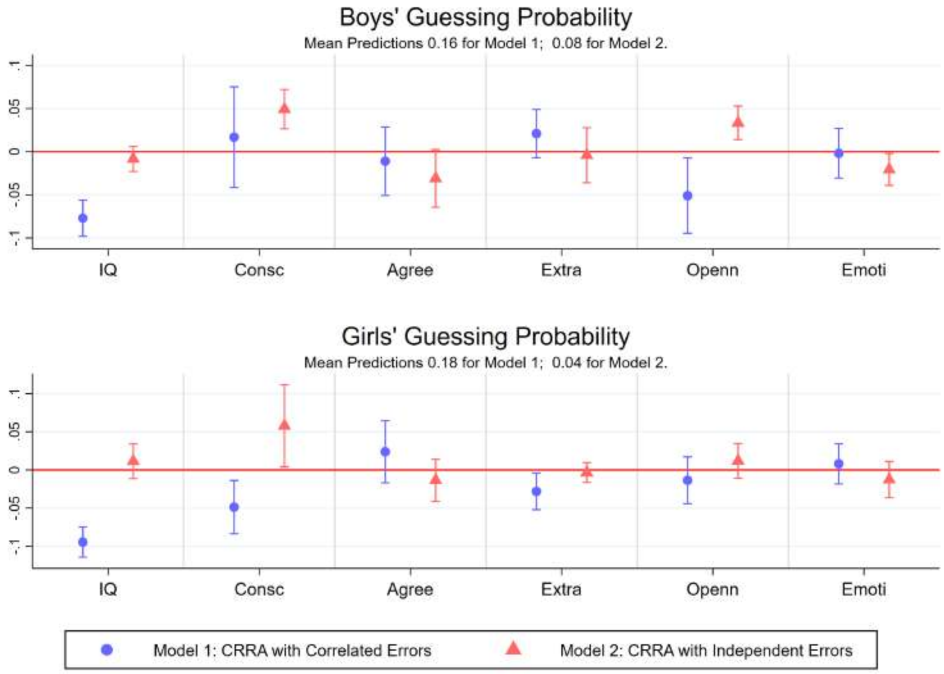


Figure L.8: The Role of ICC: Guessing Probability (CRRA)

We compute predictions for each observation and average them over the sample. Top panel: boys, bottom panel: girls. At the mean, omitting ICC underestimates guessing (boys 0.16 vs. 0.08; girls 0.18 vs. 0.04). Blue circles: Model 1 with ICC (correlated errors across items). Red triangles: Model 2 without ICC (independent errors). Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

The comparison between models with and without ICC yields three lessons. First, omitting ICC inflates the transitory noise (Expo-Power: boys 0.58 to 1.32; CRRA: boys 0.47 to 1.27). This rise is mechanical because the persistent component of error is reallocated to the transitory term. Second, omitting ICC misallocates guessing behavior to noise, underestimating the guessing probability (Expo-Power: boys 0.14 to 0.09, girls 0.19 to 0.07; CRRA: boys 0.16 to 0.08, girls 0.18 to 0.04). Third, omitting ICC underestimates risk aversion: under CRRA the risk aversion curvature r falls directly, while under Expo-Power r is essentially unchanged and the underestimation operates through the wealth-dependence parameter β. Failing to model persistent decision-noise differences therefore biases preference recovery itself, not only the noise estimates.

These patterns explain why standard models in the literature tend to find weak relationships between traits and preferences. The signal is obscured by misspecified error structure. By decomposing variance into persistent (ICC) and transitory components, we recover stronger, more interpretable trait-preference relationships while simultaneously improving model fit and reducing apparent noise. The log-likelihood improvement from adding ICC structure alone—approximately 1,960 points for boys and 1,775 for girls (Table J.1)—reflects genuine information about decision processes, not overfitting.

M Effects of Covariates Without ICC Structure

This appendix presents covariate effect estimates from specifications that include guessing behavior but exclude ICC structure. By comparing these estimates to the full model with ICC (Figures P.1–P.4), we demonstrate how ICC omission affects the estimated relationships between traits and model parameters. Models without ICC conflate persistent individual traits with transitory noise, leading to attenuated and less precise trait effects. A substantial share of apparent "noise" in standard models reflects stable individual differences rather than random decision errors, so that ICC structure is essential for recovering trait-preference relationships, not merely a statistical refinement.

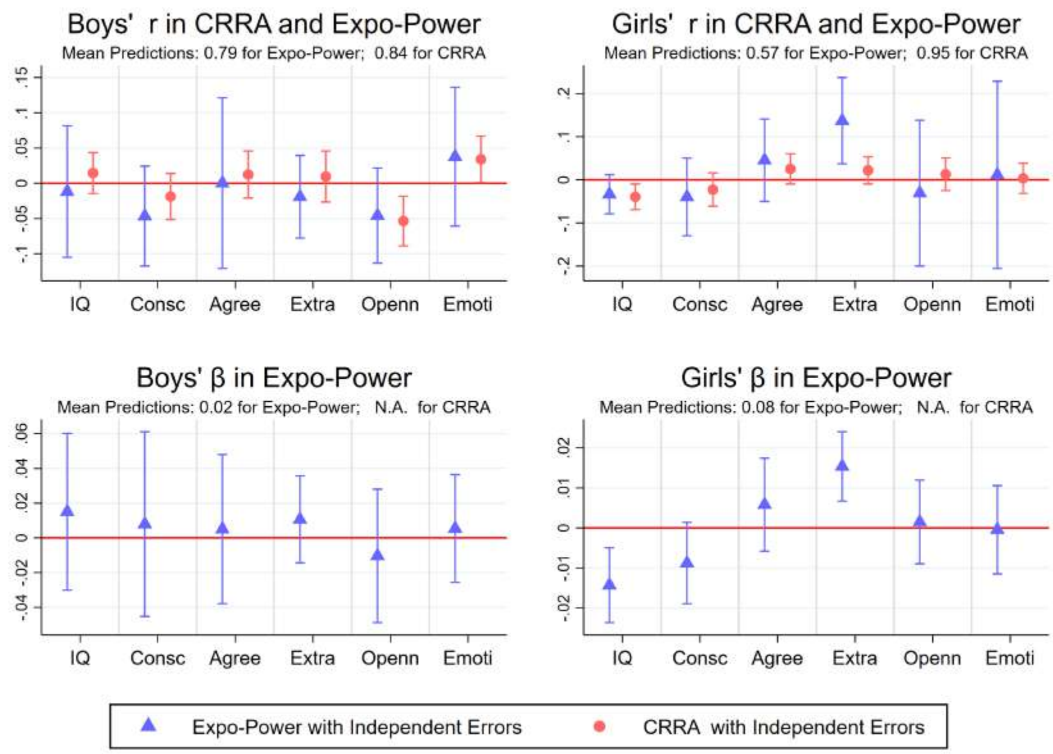


Figure M.1: Covariate Effects on Risk Aversion Parameters r and β Without ICC (Expo-Power vs. CRRA)

Notes: We compute predictions for each observation and average them over the sample. Effects of IQ and personality traits on risk aversion curvature r (top panels) and wealth-dependence β (bottom panels) in models without ICC. At the mean, omitting ICC leaves Expo-Power r close to its full-model value (boys 0.79, girls 0.57) but lowers β (boys 0.02 versus 0.09 in the full model), so risk aversion is underestimated through β; under CRRA the curvature r falls directly for boys (0.84 versus 0.97) but only marginally for girls (0.95 versus 0.97), consistent with boys' curvature being more sensitive to ICC omission. β is not defined under CRRA (N.A.). Trait effects are attenuated relative to the full model. Blue triangles: Expo-Power without ICC. Red circles: CRRA without ICC. Both specifications include guessing but exclude ICC structure. Mean predictions appear in each panel's subtitle. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

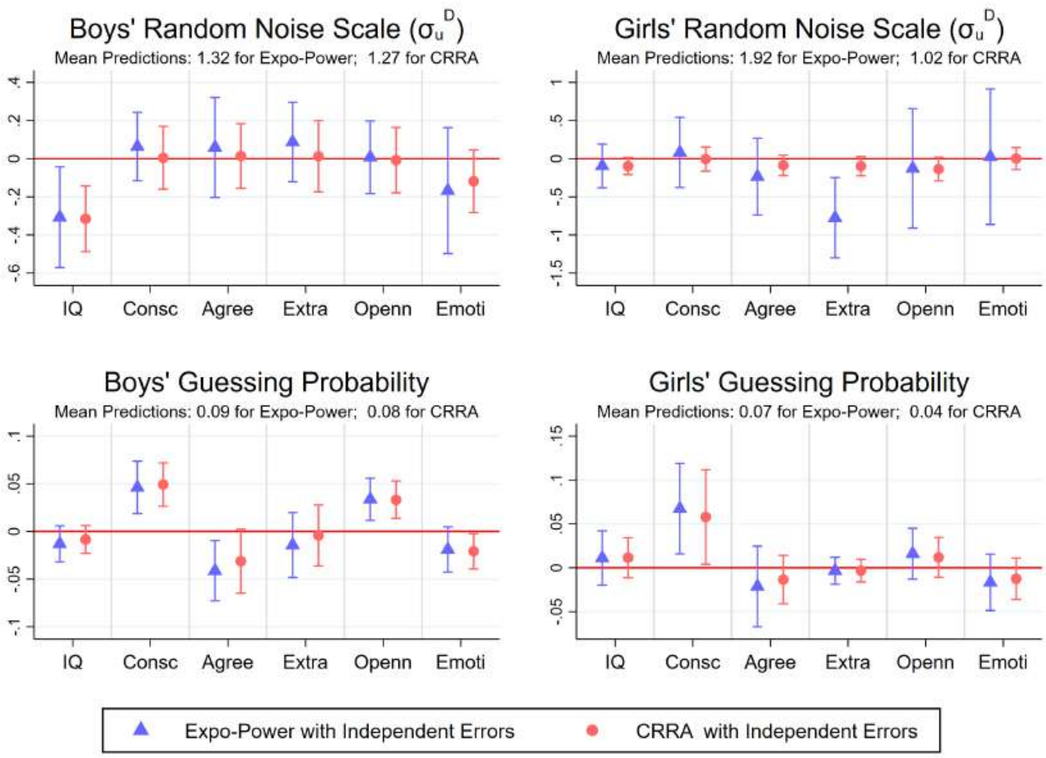


Figure M.2: Covariate Effects on Deliberation Noise Without ICC: Noise σ_{u^D} and Guessing (Expo-Power vs. CRRA)

Notes: We compute predictions for each observation and average them over the sample. Effects of IQ and personality traits on the transitory noise σ_{u^D} (top panels) and the guessing probability (bottom panels) in models without ICC. Without ICC, σ_{u^D} is inflated (boys 1.32 Expo-Power / 1.27 CRRA; girls 1.92 / 1.02) and the guessing rate is underestimated (boys 0.09 / 0.08; girls 0.07 / 0.04), as persistent individual differences are absorbed into transitory noise. Blue triangles: Expo-Power without ICC. Red circles: CRRA without ICC. Both specifications include guessing but exclude ICC structure. Mean predictions appear in each panel's subtitle. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

Two patterns stand out. First, omitting ICC inflates the transitory noise and underestimates guessing probabilities, because persistent individual differences are allocated to transitory noise. Second, omitting ICC attenuates the estimated trait effects and underestimates risk aversion—through the wealth-dependence parameter β under Expo-Power (with the curvature r essentially unchanged) and through r directly under CRRA. Failing to model persistent decision-noise differences therefore biases preference recovery itself, not just the noise estimates.

These attenuated estimates reinforce the lesson of the previous appendix. When the error structure omits persistent components, the trait-preference signal is obscured, and the large log-likelihood gains from adding ICC structure reflect genuine information about decision processes, not overfitting.

N Models Without Guessing

This appendix examines the role of explicit guessing behavior by comparing specifications that include guessing behavior (with ICC and guessing) to specifications that exclude guessing (ICC only). By removing the guessing component while retaining ICC structure, we isolate how attributing guessing behavior to deliberation noise affects parameter estimates. The comparison reveals that models without guessing roughly double the transitory noise and reduce ICC^D by more than half, reallocating guessing behavior to deliberation noise. Risk aversion is mildly underestimated—through β under Expo-Power and through r under CRRA—and, of course, the guessing-specific parameters (ICC^G, the deliberation–guessing correlation, and the guessing probability) cannot be estimated. Guessing behavior is therefore not simply "more noise" but a qualitatively distinct decision process that must be explicitly modeled for accurate preference recovery. However, the effects of eliminating guessing on structural preference parameters are small.

N.1 Expo-Power Specification

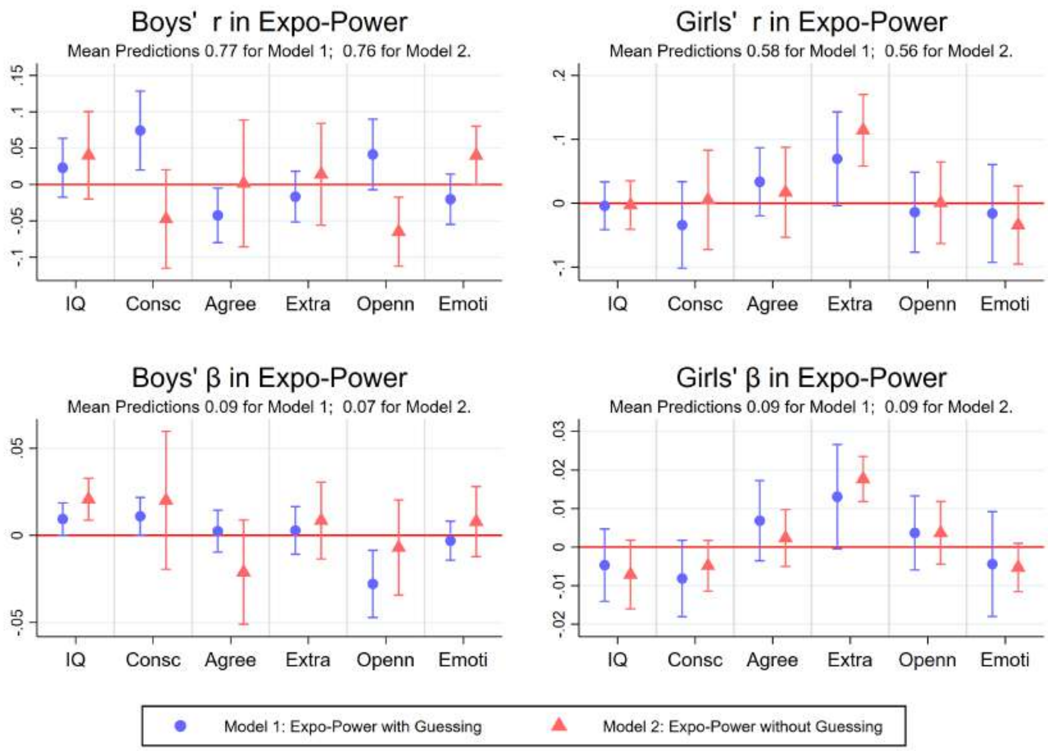


Figure N.1: The Role of Guessing: Risk Aversion Parameters r and β (Expo-Power)

We compute predictions for each observation and average them over the sample. Top panels: risk aversion curvature r. Bottom panels: wealth-dependence β. At the mean, omitting guessing leaves r essentially unchanged (boys 0.77 vs. 0.76; girls 0.58 vs. 0.56) and lowers β modestly (boys 0.09 vs. 0.07; girls 0.09 vs. 0.09), so that risk aversion is mildly underestimated, operating through β. Blue circles: Model 1 with guessing. Red triangles: Model 2 without guessing. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

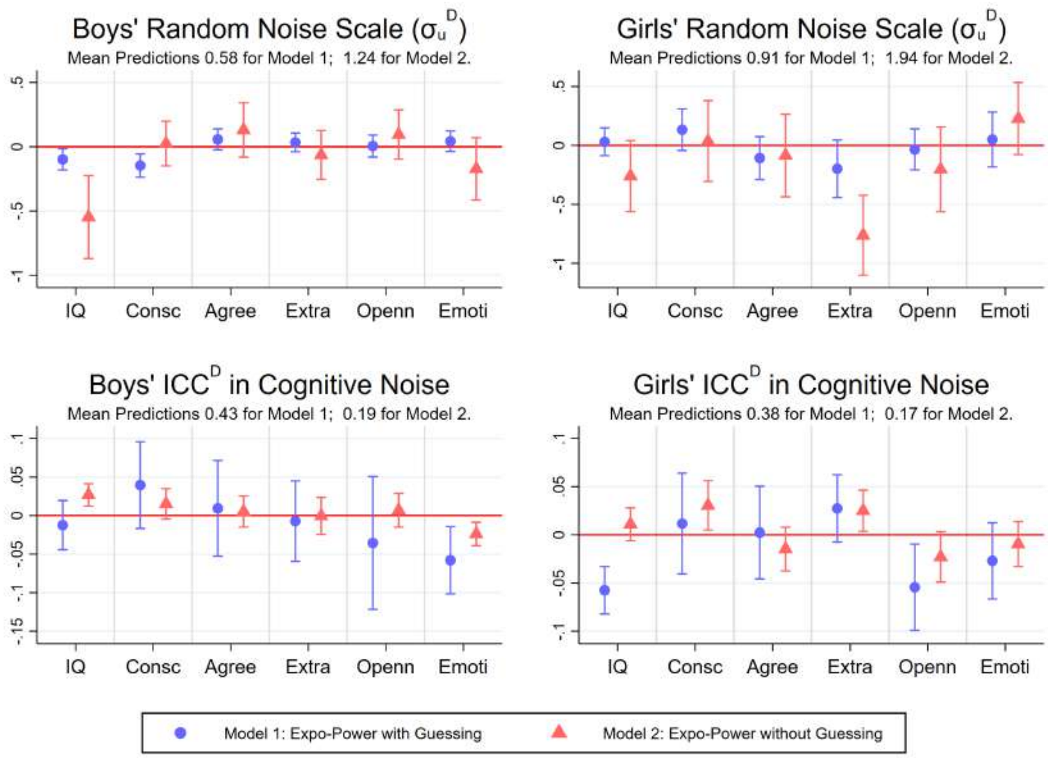


Figure N.2: The Role of Guessing: Transitory Noise σ_{u^D} and ICC^D (Expo-Power)

We compute predictions for each observation and average them over the sample. Top panels: transitory noise σ_{u^D}. Bottom panels: ICC^D. At the mean, omitting guessing inflates σ_{u^D} (boys 0.58 vs. 1.24, girls 0.91 vs. 1.94) and collapses ICC^D (boys 0.43 vs. 0.19, girls 0.38 vs. 0.17): guessing behavior is reattributed to deliberation noise, raising the transitory scale while shrinking the persistent share. Blue circles: Model 1 with guessing. Red triangles: Model 2 without guessing. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

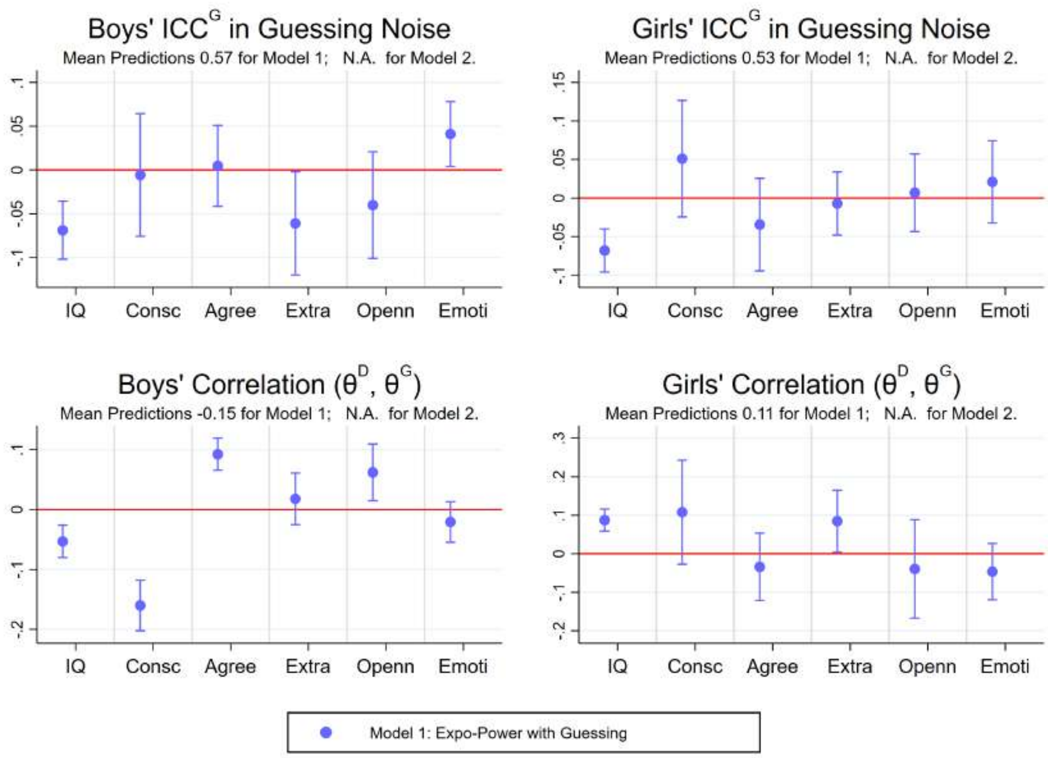


Figure N.3: The Role of Guessing: Guessing Persistence ICC^G and Deliberation–Guessing Correlation (Expo-Power)

We compute predictions for each observation and average them over the sample. Top panels: ICC^G. Bottom panels: the correlation $\rho(\theta^D, \theta^G)$. With guessing, ICC$^G = 0.57$ (boys) and 0.53 (girls), and $\rho = -0.15$ (boys) and $+0.11$ (girls). Omitting guessing discards the persistence and cross-correlation of guessing entirely. Only Model 1 (with guessing) estimates these parameters, which the model without guessing cannot (N.A.). Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

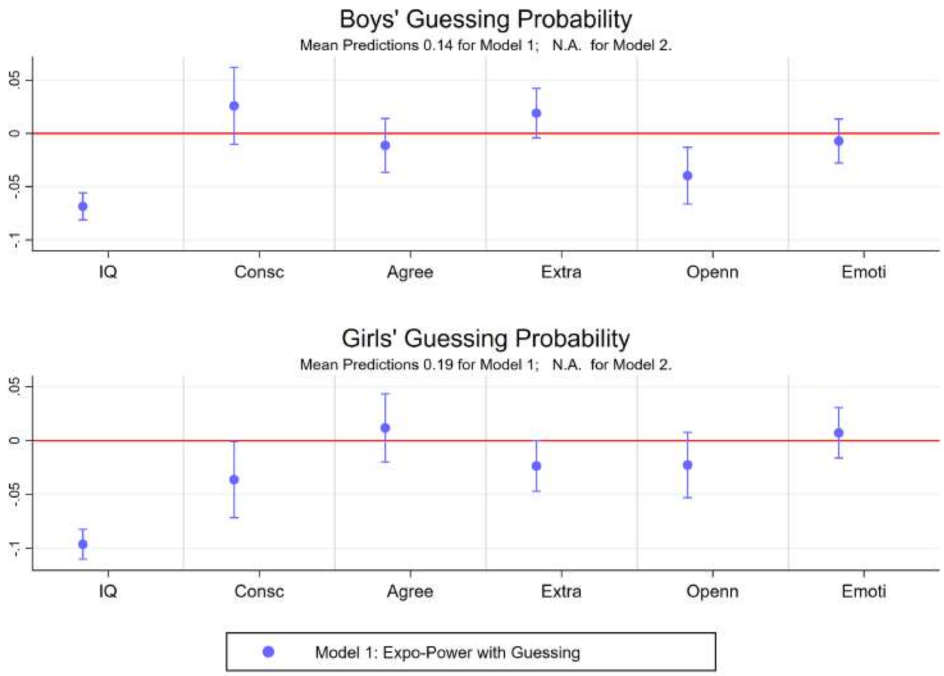


Figure N.4: The Role of Guessing: Guessing Probability (Expo-Power)

We compute predictions for each observation and average them over the sample. Effects of IQ and personality traits on the guessing probability. This parameter exists only in the model with guessing. Only Model 1 (with guessing) estimates these parameters, which the model without guessing cannot (N.A.). Error bars show 90% confidence intervals. N = 1,115 boys, N = 1,089 girls.

N.2 CRRA Specification

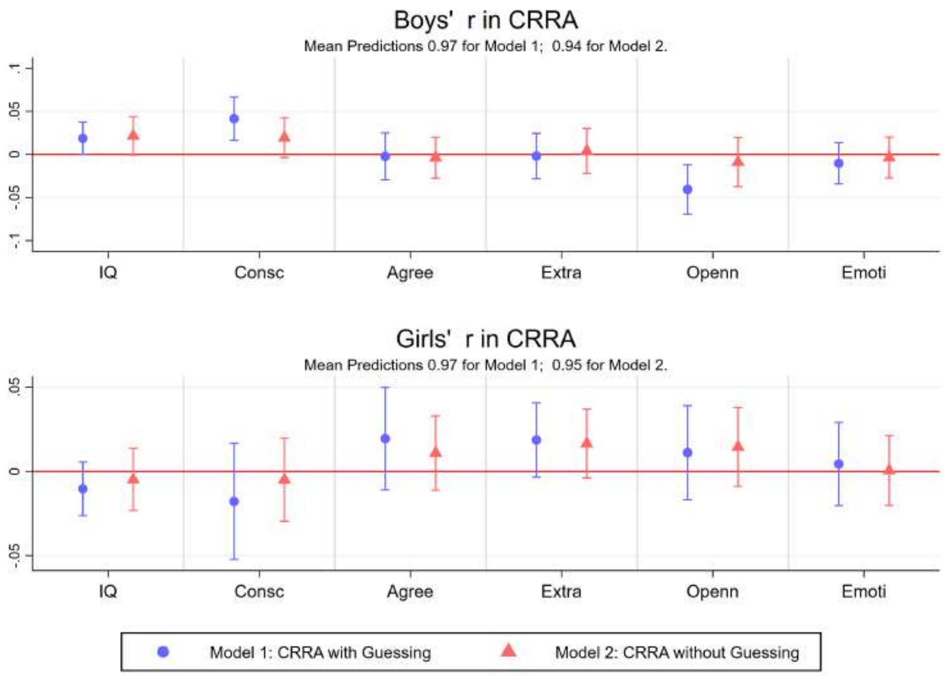


Figure N.5: The Role of Guessing: Risk Aversion r (CRRA)

We compute predictions for each observation and average them over the sample. CRRA risk aversion curvature r. Top panel: boys, bottom panel: girls. At the mean, omitting guessing lowers r slightly (boys 0.97 vs. 0.94; girls 0.97 vs. 0.95). Blue circles: Model 1 with guessing. Red triangles: Model 2 without guessing. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

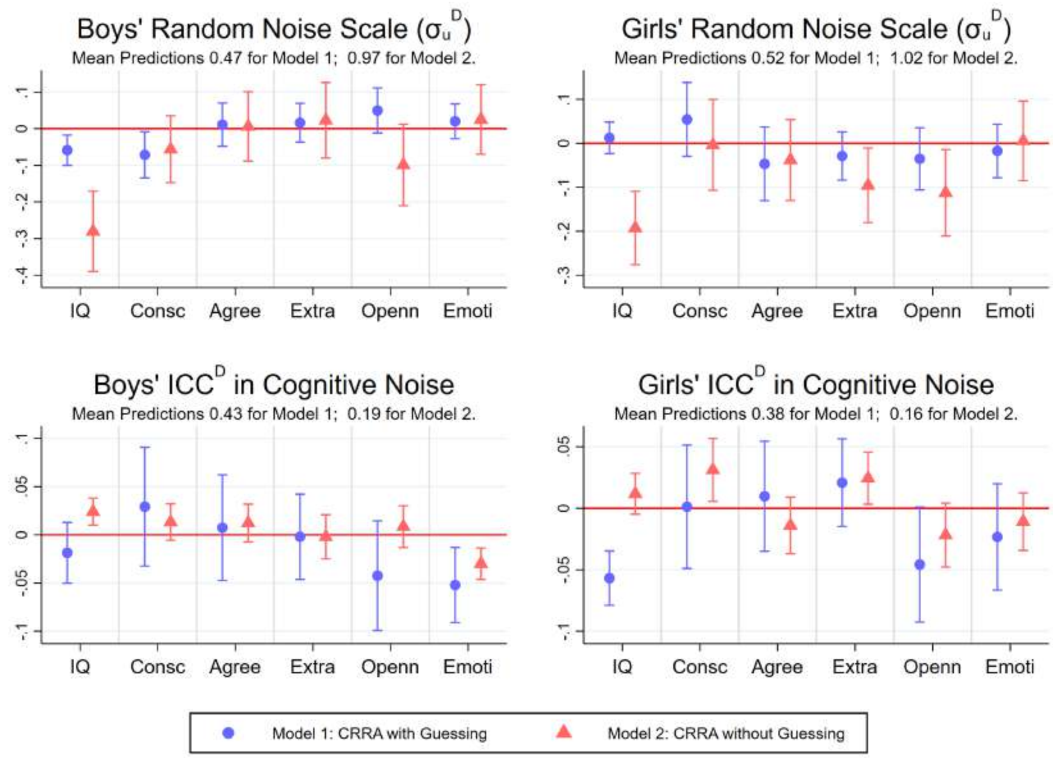


Figure N.6: The Role of Guessing: Transitory Noise σ_{u^D} and ICC^D (CRRA)

We compute predictions for each observation and average them over the sample. Top panels: σ_{u^D}. Bottom panels: ICC^D. At the mean, omitting guessing inflates σ_{u^D} (boys 0.47 vs. 0.97, girls 0.52 vs. 1.02) and collapses ICC^D (boys 0.43 vs. 0.19, girls 0.38 vs. 0.17), mirroring the Expo-Power case. Blue circles: Model 1 with guessing. Red triangles: Model 2 without guessing. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

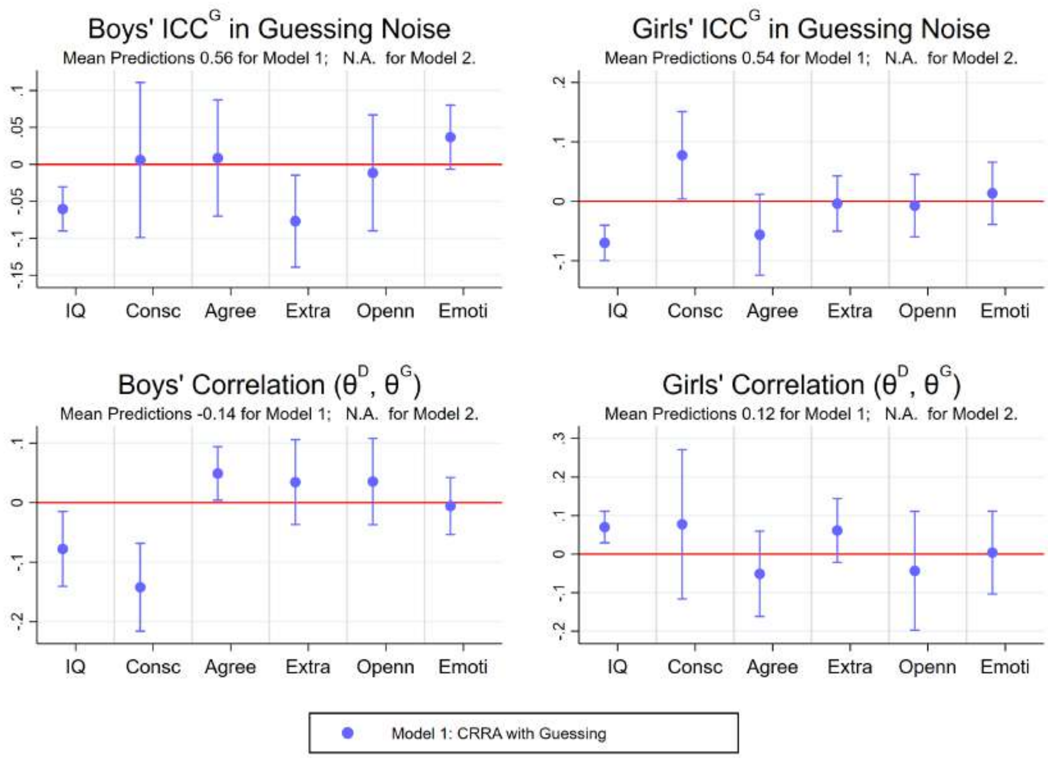


Figure N.7: The Role of Guessing: Guessing Persistence ICC^G and Deliberation–Guessing Correlation (CRRA)

We compute predictions for each observation and average them over the sample. Top panels: ICC^G. Bottom panels: $\rho(\theta^D, \theta^G)$. The values closely match the Expo-Power estimates. Only Model 1 (with guessing) estimates these parameters, which the model without guessing cannot (N.A.). Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

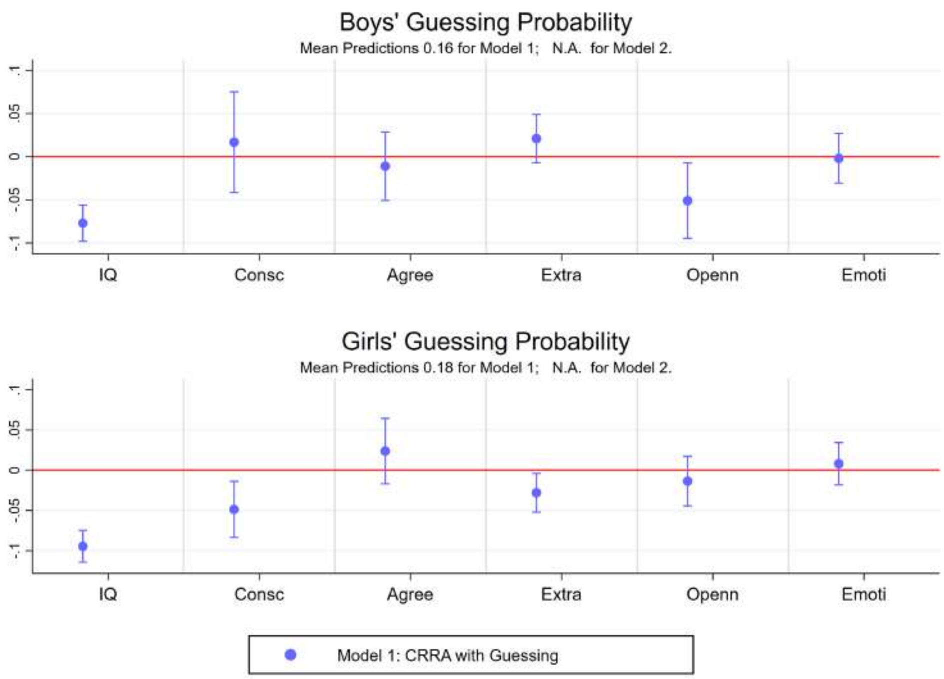


Figure N.8: The Role of Guessing: Guessing Probability (CRRA)

We compute predictions for each observation and average them over the sample. Effects of IQ and personality traits on the guessing probability under CRRA. Only Model 1 (with guessing) estimates these parameters, which the model without guessing cannot (N.A.). Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

O Full Model Coefficient Tables

This appendix presents detailed coefficient estimates for all model parameters in the full specification with ICC structure, guessing behavior, and trait-based heterogeneity. Table O.1 reports effects on risk preferences and deliberation noise. Table O.2 reports effects on ICC structure, error correlations, and guessing probability. Graphical summaries appear in Appendix P.

Table O.1: Full Model: Effects of IQ and Personality Traits on Risk Preferences and Deliberation Noise

	Expo-Power		CRRA	Deliberation Noise	
Variable	$r(\boldsymbol{X})$	$\beta(\boldsymbol{X})$	$r(\boldsymbol{X})$	$\sigma_{u^D}(\boldsymbol{X})$ (Expo-Power)	$\sigma_{u^D}(\boldsymbol{X})$ (CRRA)
Panel A: Boys					
Constant	0.771^{***}	0.091^{***}	0.970^{***}	0.577^{***}	0.469^{***}
	(0.033)	(0.009)	(0.012)	(0.046)	(0.028)
IQ	0.023	0.009^{*}	0.019	-0.099^{*}	-0.059^{**}
	(0.025)	(0.006)	(0.011)	(0.050)	(0.025)
Conscientiousness	0.074^{**}	0.011^{*}	0.041^{***}	-0.147^{***}	-0.071^{*}
	(0.033)	(0.007)	(0.015)	(0.055)	(0.038)
Agreeableness	-0.042^{*}	0.002	-0.002	0.057	0.011
	(0.023)	(0.007)	(0.017)	(0.049)	(0.036)
Extraversion	-0.017	0.003	-0.002	0.034	0.016
	(0.021)	(0.008)	(0.016)	(0.044)	(0.032)
Openness	0.041	-0.028^{**}	-0.041^{**}	0.005	0.050
	(0.030)	(0.012)	(0.017)	(0.052)	(0.037)
Emotional Stability	-0.020	-0.003	-0.010	0.043	0.020
	(0.021)	(0.007)	(0.015)	(0.048)	(0.029)
Panel B: Girls					
Constant	0.576^{***}	0.093^{***}	0.968^{***}	0.908^{***}	0.519^{***}
	(0.025)	(0.005)	(0.011)	(0.082)	(0.028)
IQ	-0.004	-0.005	-0.010	0.031	0.013
	(0.023)	(0.006)	(0.010)	(0.072)	(0.022)
Conscientiousness	-0.034	-0.008	-0.018	0.134	0.054
	(0.041)	(0.006)	(0.021)	(0.107)	(0.051)
Agreeableness	0.034	0.007	0.019	-0.107	-0.047
	(0.032)	(0.006)	(0.019)	(0.111)	(0.051)
Extraversion	0.070	0.013	0.019	-0.198	-0.029
	(0.045)	(0.008)	(0.013)	(0.148)	(0.033)
Openness	-0.014	0.004	0.011	-0.034	-0.035
	(0.038)	(0.006)	(0.017)	(0.106)	(0.043)
Emotional Stability	-0.016	-0.004	0.004	0.051	-0.017
	(0.047)	(0.008)	(0.015)	(0.141)	(0.037)

Notes: Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. This table presents coefficient estimates from the full model specification including ICC structure and guessing behavior. All models include IQ and Big Five personality traits as covariates. σ_{u^D} represents the standard deviation of transitory (random) noise in the deliberation equation. Graphical representations of these coefficients appear in Figures P.1 and P.2. $N = 1{,}115$ for boys and $N = 1{,}089$ for girls.

Table O.2: Full Model: Effects of IQ and Personality Traits on ICC, Correlation, and Guessing

	Intra-Class Correlation				Correlation and Guessing			
	Expo-Power		CRRA		Expo-Power		CRRA	
Variable	$\text{ICC}^D(\boldsymbol{X})$	$\text{ICC}^G(\boldsymbol{X})$	$\text{ICC}^D(\boldsymbol{X})$	$\text{ICC}^G(\boldsymbol{X})$	$\text{Corr}(\theta^D,\theta^G)(\boldsymbol{X})$	$\text{Pr(Guess)}(\boldsymbol{X})$	$\text{Corr}(\theta^D,\theta^G)(\boldsymbol{X})$	$\text{Pr(Guess)}(\boldsymbol{X})$
Panel A: Boys								
Constant	0.428***	0.574***	0.433***	0.563***	−0.150***	0.144***	−0.142***	0.158***
	(0.033)	(0.025)	(0.025)	(0.024)	(0.034)	(0.014)	(0.039)	(0.019)
IQ	−0.012	−0.069***	−0.019	−0.060***	−0.053***	−0.069***	−0.078**	−0.077***
	(0.019)	(0.020)	(0.019)	(0.018)	(0.016)	(0.008)	(0.038)	(0.013)
Conscientiousness	0.039	−0.006	0.029	0.006	−0.160***	0.026	−0.142***	0.017
	(0.034)	(0.043)	(0.037)	(0.064)	(0.026)	(0.022)	(0.045)	(0.035)
Agreeableness	0.009	0.005	0.007	0.009	0.092***	−0.011	0.049*	−0.011
	(0.038)	(0.028)	(0.033)	(0.048)	(0.016)	(0.015)	(0.027)	(0.024)
Extraversion	−0.007	−0.061*	−0.002	−0.077**	0.018	0.019	0.035	0.021
	(0.032)	(0.036)	(0.027)	(0.038)	(0.026)	(0.014)	(0.043)	(0.017)
Openness	−0.036	−0.040	−0.042	−0.012	0.062**	−0.040**	0.036	−0.051*
	(0.052)	(0.037)	(0.035)	(0.048)	(0.029)	(0.016)	(0.044)	(0.027)
Emotional Stability	−0.058**	0.041*	−0.052**	0.037	−0.021	−0.007	−0.006	−0.002
	(0.027)	(0.023)	(0.024)	(0.026)	(0.021)	(0.013)	(0.029)	(0.018)
Panel B: Girls								
Constant	0.383***	0.531***	0.376***	0.540***	0.108**	0.187***	0.125**	0.184***
	(0.019)	(0.023)	(0.019)	(0.024)	(0.045)	(0.012)	(0.056)	(0.015)
IQ	−0.057***	−0.068***	−0.057***	−0.070***	0.087***	−0.096***	0.070***	−0.095***
	(0.015)	(0.017)	(0.013)	(0.018)	(0.018)	(0.008)	(0.025)	(0.012)
Conscientiousness	0.012	0.051	0.001	0.078*	0.108	−0.036*	0.078	−0.049**
	(0.032)	(0.046)	(0.030)	(0.045)	(0.082)	(0.021)	(0.118)	(0.021)
Agreeableness	0.002	−0.034	0.010	−0.056	−0.034	0.012	−0.051	0.024
	(0.029)	(0.037)	(0.027)	(0.041)	(0.053)	(0.019)	(0.067)	(0.025)
Extraversion	0.027	−0.007	0.021	−0.004	0.085*	−0.024*	0.061	−0.028*
	(0.021)	(0.025)	(0.022)	(0.028)	(0.049)	(0.014)	(0.050)	(0.015)
Openness	−0.054**	0.007	−0.046	−0.007	−0.039	−0.023	−0.043	−0.014
	(0.027)	(0.031)	(0.028)	(0.032)	(0.078)	(0.018)	(0.094)	(0.019)
Emotional Stability	−0.027	0.021	−0.023	0.014	−0.046	0.007	0.004	0.008
	(0.024)	(0.032)	(0.026)	(0.032)	(0.044)	(0.014)	(0.065)	(0.016)

Notes: Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. This table presents coefficient estimates from the full model specification. ICC^D measures the intra-class correlation in the deliberation equation, capturing persistent individual differences in deliberation noise. ICC^G measures the intra-class correlation in the guessing equation, capturing persistent individual differences in guessing. $\rho(\theta^D, \theta^G)$ is the correlation between deliberation and guessing random effects. Pr(Guess) is the probability of guessing. The correlation and guessing-probability parameters are reported separately for the Expo-Power and CRRA specifications. Graphical representations of these coefficients appear in Figures P.2, P.3, and P.4. $N = 1{,}115$ for boys and $N = 1{,}089$ for girls.

O.1 Predicted Values Across Specifications

Tables O.3 and O.4 report the predicted values over the sampling distribution of IQ and traits across all specifications for both the Expo-Power and CRRA models. Table 5 in the main text reports the Expo-Power rows only. Risk aversion estimates are consistently higher under CRRA than Expo-Power, with the gap approximately 0.20 in risk aversion curvature in the full specification for boys ($r = 0.970$ vs. 0.771) and approximately 0.39

for girls ($r = 0.968$ vs. 0.576), reflecting the binding nature of CRRA's functional form restrictions, especially for girls.

Table O.3: Predicted Values of Model Parameters over Sampling Distribution of IQ and Traits: Boys (Expo-Power and CRRA)

	Parameter						
	$r(\boldsymbol{X})$	$\beta(\boldsymbol{X})$	$\sigma_{u^D}(\boldsymbol{X})$	ICC$^D(\boldsymbol{X})$	ICC$^G(\boldsymbol{X})$	Corr$(\theta^D,\theta^G)(\boldsymbol{X})$	Pr(Guess)$(\boldsymbol{X})$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Expo-Power, with IQ and personality traits							
With ICC and Guessing	0.771***	0.091***	0.577***	0.428***	0.574***	−0.150***	0.144***
	(0.033)	(0.009)	(0.046)	(0.033)	(0.025)	(0.034)	(0.014)
No ICC, with Guessing	0.793***	0.018	1.320***				0.090***
	(0.161)	(0.072)	(0.267)				(0.023)
With ICC, No Guessing	0.765***	0.074***	1.241***	0.190***			
	(0.097)	(0.027)	(0.226)	(0.010)			
No ICC, No Guessing	0.698***	0.043**	1.871***				
	(0.091)	(0.017)	(0.462)				
CRRA, with IQ and personality traits							
With ICC and Guessing	0.970***		0.469***	0.433***	0.563***	−0.142***	0.158***
	(0.012)		(0.028)	(0.025)	(0.024)	(0.039)	(0.019)
No ICC, with Guessing	0.840***		1.269***				0.083***
	(0.015)		(0.101)				(0.021)
With ICC, No Guessing	0.936***		0.971***	0.187***			
	(0.011)		(0.057)	(0.010)			
No ICC, No Guessing	0.842***		1.458***				
	(0.015)		(0.099)				
Expo-Power, without IQ or personality traits							
With ICC and Guessing	0.777***	0.099***	0.519***	0.425***	0.632***	−0.189***	0.126***
	(0.029)	(0.007)	(0.034)	(0.020)	(0.023)	(0.044)	(0.012)
No ICC, with Guessing	0.711***	0.047***	1.227***				0.164***
	(0.041)	(0.009)	(0.098)				(0.018)
With ICC, No Guessing	0.760***	0.082***	1.046***	0.180***			
	(0.034)	(0.007)	(0.072)	(0.010)			
No ICC, No Guessing	0.767***	0.034*	1.485***				
	(0.055)	(0.018)	(0.128)				
CRRA, without IQ or personality traits							
With ICC and Guessing	0.975***		0.446***	0.422***	0.626***	−0.177***	0.127***
	(0.012)		(0.022)	(0.020)	(0.023)	(0.043)	(0.012)
No ICC, with Guessing	0.843***		1.062***				0.156***
	(0.015)		(0.066)				(0.018)
With ICC, No Guessing	0.939***		0.895***	0.178***			
	(0.010)		(0.038)	(0.010)			
No ICC, No Guessing	0.845***		1.368***				
	(0.014)		(0.073)				

Notes: We compute predictions for each observation and average them over the sample. Each entry is the prediction over the sampling distribution of IQ and traits. Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The parameters are: r, the risk aversion curvature of the utility function (for CRRA, the single risk-aversion parameter; for Expo-Power, the risk aversion curvature, with $\beta = 0$ by construction under CRRA); β, the wealth-dependence of relative risk aversion; σ_{u^D}, the standard deviation of transitory noise; ICC^D and ICC^G, the intra-class correlations of the persistent components of deliberation and guessing noise; $Corr(\theta^D, \theta^G)$, the correlation between those persistent components; and Pr(Guess), the mean probability of guessing. $N = 1{,}115$.

Table O.4: Predicted Values of Model Parameters over Sampling Distribution of IQ and Traits: Girls (Expo-Power and CRRA)

	Parameter						
	$r(\boldsymbol{X})$	$\beta(\boldsymbol{X})$	$\sigma_{u^D}(\boldsymbol{X})$	$\text{ICC}^D(\boldsymbol{X})$	$\text{ICC}^G(\boldsymbol{X})$	$\text{Corr}(\theta^D,\theta^G)(\boldsymbol{X})$	$\text{Pr(Guess)}(\boldsymbol{X})$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Expo-Power, with IQ and personality traits							
With ICC and Guessing	0.576***	0.093***	0.908***	0.383***	0.531***	0.108**	0.187***
	(0.025)	(0.005)	(0.082)	(0.019)	(0.023)	(0.045)	(0.012)
No ICC, with Guessing	0.573***	0.083***	1.916***				0.073***
	(0.069)	(0.005)	(0.282)				(0.025)
With ICC, No Guessing	0.557***	0.087***	1.941***	0.166***			
	(0.030)	(0.004)	(0.171)	(0.009)			
No ICC, No Guessing	0.579***	0.080***	2.178***				
	(0.037)	(0.005)	(0.216)				
CRRA, with IQ and personality traits							
With ICC and Guessing	0.968***		0.519***	0.376***	0.540***	0.125**	0.184***
	(0.011)		(0.028)	(0.019)	(0.024)	(0.056)	(0.015)
No ICC, with Guessing	0.953***		1.016***				0.044*
	(0.014)		(0.069)				(0.024)
With ICC, No Guessing	0.949***		1.016***	0.165***			
	(0.009)		(0.044)	(0.009)			
No ICC, No Guessing	0.950***		1.081***				
	(0.014)		(0.060)				
Expo-Power, without IQ or personality traits							
With ICC and Guessing	0.582***	0.095***	0.850***	0.384***	0.575***	0.056	0.164***
	(0.023)	(0.005)	(0.062)	(0.021)	(0.023)	(0.084)	(0.014)
No ICC, with Guessing	0.527***	0.078***	1.678***				0.143***
	(0.024)	(0.004)	(0.141)				(0.024)
With ICC, No Guessing	0.553***	0.086***	1.720***	0.163***			
	(0.024)	(0.004)	(0.134)	(0.009)			
No ICC, No Guessing	0.546***	0.080***	2.013***				
	(0.025)	(0.004)	(0.163)				
CRRA, without IQ or personality traits							
With ICC and Guessing	0.972***		0.500***	0.381***	0.582***	0.078	0.163***
	(0.011)		(0.024)	(0.021)	(0.021)	(0.070)	(0.014)
No ICC, with Guessing	0.948***		0.956***				0.067**
	(0.014)		(0.075)				(0.033)
With ICC, No Guessing	0.949***		0.965***	0.161***			
	(0.009)		(0.037)	(0.009)			
No ICC, No Guessing	0.946***		1.056***				
	(0.013)		(0.053)				

Notes: We compute predictions for each observation and average them over the sample. Each entry is the prediction over the sampling distribution of IQ and traits. Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The parameters are defined as in Table 5 in the main text: r is the risk aversion curvature ($\beta = 0$ by construction under CRRA), β the wealth-dependence of relative risk aversion, σ_{u^D} the transitory noise, ICC^D and ICC^G the intra-class correlations of the persistent deliberation and guessing components, $Corr(\theta^D, \theta^G)$ their correlation, and Pr(Guess) the mean probability of guessing. $N = 1{,}089$.

O.2 Joint Significance and Gender-Difference Tests

Table O.5 reports the Wald tests for the joint significance of the personality traits on each full-model parameter, and Table O.6 reports the gender-difference decomposition, both for the Expo-Power and CRRA specifications. Tables 6 and 7 in the main text report the Expo-Power rows only.

Table O.5: Joint Significance Tests: Personality Trait Effects on Model Parameters (Expo-Power and CRRA)

	Parameters							
	$r(\boldsymbol{X})$	$\beta(\boldsymbol{X})$	$\sigma_{u^D}(\boldsymbol{X})$	$ICC^D(\boldsymbol{X})$	$ICC^G(\boldsymbol{X})$	Corr$(\theta^D, \theta^G)(\boldsymbol{X})$	Pr(Guess)$(\boldsymbol{X})$	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Boys, Expo-Power								
With ICC and Guessing	0.028	0.055	0.072	0.012	0.003	0.000	0.021	0.000
No ICC, with Guessing	0.032	0.705	0.903				0.008	0.000
With ICC, No Guessing	0.010	0.094	0.771	0.155				0.000
No ICC, No Guessing	0.010	0.016	0.569					0.000
Boys, CRRA								
With ICC and Guessing	0.010		0.057	0.033	0.014	0.004	0.005	0.000
No ICC, with Guessing	0.096		0.852				0.006	0.000
With ICC, No Guessing	0.841		0.189	0.064				0.000
No ICC, No Guessing	0.119		0.559					0.000
Girls, Expo-Power								
With ICC and Guessing	0.496	0.053	0.247	0.329	0.607	0.000	0.006	0.000
No ICC, with Guessing	0.147	0.000	0.001				0.081	0.000
With ICC, No Guessing	0.000	0.000	0.000	0.175				0.000
No ICC, No Guessing	0.000	0.000	0.000					0.000
Girls, CRRA								
With ICC and Guessing	0.169		0.345	0.519	0.505	0.466	0.008	0.000
No ICC, with Guessing	0.293		0.010				0.032	0.000
With ICC, No Guessing	0.070		0.000	0.162				0.000
No ICC, No Guessing	0.263		0.014					0.000

Notes: Each entry is a p-value for the joint null hypothesis that the five personality traits have no effect on the indicated parameter. Wald tests with 5 degrees of freedom. IQ enters the model but is not included in this joint test. Values below 0.05 indicate rejection at the 5% level. Empty cells indicate parameters not estimated in that specification. Column (8) tests whether all trait coefficients across all parameters are jointly zero.

Table O.6: Sources of Girl–Boy Differences in Parameter Estimates Evaluated at Mean Boy $\bar{\boldsymbol{X}}_B$

	Parameters						
	$r(\boldsymbol{X})$	$\beta(\boldsymbol{X})$	$\sigma_{u^D}(\boldsymbol{X})$	$ICC^D(\boldsymbol{X})$	$ICC^G(\boldsymbol{X})$	$\text{Corr}(\theta^D, \theta^G)(\boldsymbol{X})$	$\text{Pr(Guess)}(\boldsymbol{X})$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Expo-Power, With ICC and Guessing							
Total	-0.196***	0.002	0.251***	-0.050	-0.046	0.267***	0.030
Term 1 (Composition)	0.025*	0.006*	-0.051***	0.009	0.006	-0.066***	0.006
Term 2 (Coefficient)	-0.183**	0.005	0.217	-0.047	-0.071	0.226	0.048
Term 3 (Interaction)	-0.039	-0.009**	0.085**	-0.012	0.019	0.107	-0.024**
Expo-Power, No ICC, with Guessing							
Total	-0.084	0.032***	0.200				-0.063
Term 1 (Composition)	0.020	-0.001	-0.012				-0.016**
Term 2 (Coefficient)	-0.077	0.035	0.218				-0.101
Term 3 (Interaction)	-0.027	-0.002	-0.006				0.053**
Expo-Power, With ICC, No Guessing							
Total	-0.153***	0.007	0.422***	-0.023			
Term 1 (Composition)	0.026*	0.000	-0.076	0.005			
Term 2 (Coefficient)	-0.152*	0.011	0.398	-0.034			
Term 3 (Interaction)	-0.027	-0.004	0.100	0.006			
Expo-Power, No ICC, No Guessing							
Total	-0.120	0.038**	0.181				
Term 1 (Composition)	0.025**	-0.003	-0.082				
Term 2 (Coefficient)	-0.120	0.041	0.157				
Term 3 (Interaction)	-0.025	0.000	0.107				
CRRA, With ICC and Guessing							
Total	-0.001		0.046	-0.059*	-0.024	0.261***	0.026
Term 1 (Composition)	0.019***		-0.029**	0.003	0.014	-0.054**	-0.002
Term 2 (Coefficient)	0.002		0.038	-0.054	-0.048	0.235	0.043
Term 3 (Interaction)	-0.021**		0.037*	-0.008	0.009	0.080*	-0.015
CRRA, No ICC, with Guessing							
Total	0.112***		-0.265**				-0.070
Term 1 (Composition)	0.001		-0.029				0.013*
Term 2 (Coefficient)	0.117***		-0.224				-0.114
Term 3 (Interaction)	-0.006		-0.011				0.031
CRRA, With ICC, No Guessing							
Total	0.013		0.060	-0.023			
Term 1 (Composition)	0.009		-0.030	0.005			
Term 2 (Coefficient)	0.012		0.077	-0.034			
Term 3 (Interaction)	-0.008		0.013	0.006			
CRRA, No ICC, No Guessing							
Total	0.108***		-0.388***				
Term 1 (Composition)	0.002		-0.011				
Term 2 (Coefficient)	0.114***		-0.402*				
Term 3 (Interaction)	-0.009		0.024				

Notes: For each parameter and specification, we report the total girl–boy difference in predicted values together with Terms 1, 2, and 3 from Equation (20). The total is $F^q(\bar{\boldsymbol{X}}_G, \boldsymbol{b}_G^q) - F^q(\bar{\boldsymbol{X}}_B, \boldsymbol{b}_B^q)$, the difference in the predicted column parameter when each gender is evaluated at its own mean IQ and personality traits under its own coefficients. Term 1 (composition, endowments) is $\frac{\partial F^q}{\partial \boldsymbol{X}^g}(\bar{\boldsymbol{X}}_G - \bar{\boldsymbol{X}}_B)$, which holds coefficients at the boys' values and varies the mean covariates. Term 2 (coefficient, structure) is $\frac{\partial F^q}{\partial \boldsymbol{b}^q}(\boldsymbol{b}_G^q - \boldsymbol{b}_B^q)$, which holds covariates at the boys' means and varies the coefficients. Term 3 (interaction) is $\frac{\partial^2 F^q}{\partial \boldsymbol{b}^q\, \partial \boldsymbol{X}^g}(\bar{\boldsymbol{X}}_G - \bar{\boldsymbol{X}}_B)(\boldsymbol{b}_G^q - \boldsymbol{b}_B^q)$, the part that depends jointly on both differences. All derivatives are evaluated at the boys' means and coefficients $(\bar{\boldsymbol{X}}_B, \boldsymbol{b}_B^q)$. Boys are the reference group. Linear decompositions of parameters involving nonlinear transformations are approximated using these differentials for continuous covariates and discrete differences for binary covariates, evaluated at the sample means. Standard errors are computed by the delta method. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Empty cells indicate parameters not estimated in that specification. Because these totals are evaluated at mean trait levels rather than averaged over the sample, they can differ modestly from average-over-sample parameter values.

The decomposition sharpens the contrast between the two functional forms. Under Expo-Power with ICC and guessing, girls have substantially lower risk aversion curvature r than boys (total difference -0.196, $p < 0.01$), and this gap is structural: the coefficient component alone accounts for -0.183 ($p < 0.05$), while the composition component is small and of the opposite sign (0.025, $p < 0.1$). Under CRRA, by contrast, the total gender difference in r is essentially zero and statistically insignificant (-0.001). Its near-zero magnitude reflects offsetting decomposition components—a significantly positive composition term (0.019, $p < 0.01$) cancelled by a significantly negative interaction term (-0.021, $p < 0.05$), with a negligible and insignificant coefficient term (0.002)—rather than an absence of underlying differences. This illustrates how CRRA's functional-form restriction compresses the estimated gender gap in risk aversion toward zero, whereas the more flexible Expo-Power specification reveals a large, structurally driven difference.

The persistent components of the error structure show little gender difference. Under CRRA, the gap in ICC^G is small and insignificant (-0.024), and none of its decomposition components is significant; the gap in ICC^D is marginally significant in total (-0.059, $p < 0.1$) but is not sharply attributable to any single component, since none of its components reaches significance individually. This suggests that the proportion of decision variance attributable to stable traits is broadly similar across genders. Transitory noise σ_{u^D} behaves differently across functional forms: under Expo-Power girls are significantly noisier than boys (total 0.251, $p < 0.01$), whereas under CRRA the gender difference is small and insignificant (0.046). For the Expo-Power gap, the coefficient component is the largest in magnitude (0.217) though not individually significant, the composition component is significantly negative (-0.051, $p < 0.01$), and the interaction component is significantly positive (0.085, $p < 0.05$).

The correlation between the deliberation and guessing components, $\rho(\theta^D, \theta^G)$, shows a large and highly significant total gender difference under both functional forms (Expo-Power 0.267, CRRA 0.261; both $p < 0.01$). This reflects correlations of opposite sign:

negative for boys ($\rho \approx -0.14$ to -0.15) and positive for girls ($\rho \approx +0.11$ to $+0.13$). Within the CRRA decomposition, the coefficient component is the largest in magnitude (0.235) though not individually significant, while the composition component is significantly negative (-0.054, $p < 0.05$); the sign reversal therefore does not arise merely from gender differences in trait endowments, and instead points to a genuinely different relationship between deliberation precision and guessing propensity across genders.

The decomposition also clarifies how to read the trait-by-trait coefficient plots. For example, Conscientiousness significantly raises r for boys (0.074, $p < 0.05$ under Expo-Power) but not for girls (-0.034, insignificant), and this is not a mechanical consequence of boys and girls differing in mean Conscientiousness: the coefficient difference itself is what the decomposition isolates. More generally, the economic consequences of a given trait—how much a one-unit change shifts a parameter—differ by gender, so the structural (coefficient) components are not simply compositional artifacts of girls scoring higher on Conscientiousness or Agreeableness. This supports the broader conclusion that boys and girls exhibit distinct patterns in how cognitive and personality traits map into economic preferences.

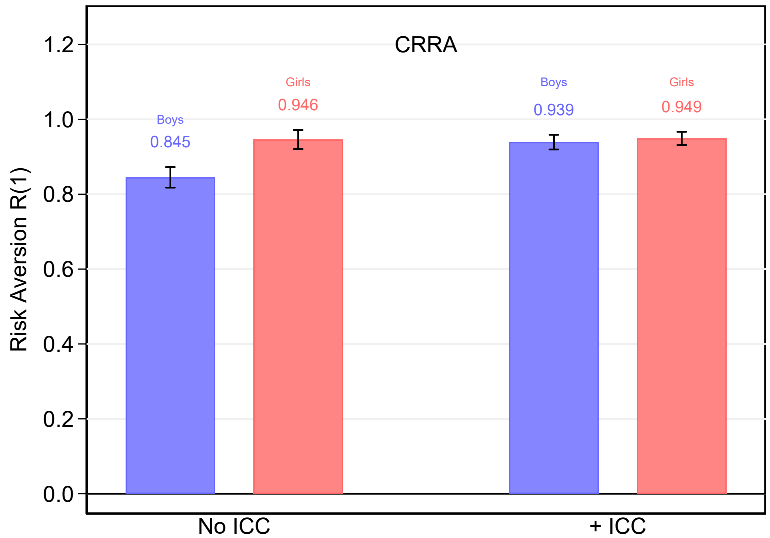


Figure O.1: Gender Gaps in Risk Aversion under CRRA

Notes: Risk aversion $\text{RRA}(s+w) = r$ at $w = 1$ by gender under CRRA, without versus with ICC structure, from Table J.1. Without ICC, girls appear more risk-averse than boys (0.946 vs. 0.845), replicating the conventional finding. Adding ICC narrows the gap to near zero (boys 0.939, girls 0.949), so that correlated deliberation noise—not preferences—drives the apparent gender difference. Blue bars: boys, red bars: girls. Error bars represent 90% confidence intervals.

P Effects of IQ and Personality Traits on Risk Preferences, Deliberation Noise, and Guessing

P.1 Risk Preferences

Figure P.1 presents the effects of IQ and personality traits on overall risk aversion $\text{RRA}(s+w) = r + \beta(s+w)^{1-r}$ evaluated at $w = 1$. To compute this, we evaluate their impacts on r and β. The figure compares Expo-Power (blue triangles) and CRRA (red circles) specifications.

For boys, Conscientiousness significantly increases risk aversion curvature r under both specifications (Expo-Power: 0.074, $p < 0.05$, CRRA: 0.041, $p < 0.01$). This finding aligns with theoretical predictions: conscientious individuals, characterized by deliberation and impulse control, may exhibit greater caution in risky choices. Agreeableness shows a marginally negative effect under Expo-Power (-0.042, $p < 0.10$), suggesting more agreeable boys are slightly less risk-averse.

For girls, the pattern differs substantially. No trait reaches statistical significance on the preference parameters. The largest point estimate is for Extraversion (Expo-Power: 0.070, SE 0.045, CRRA: 0.019, SE 0.013), but neither effect is statistically significant. IQ shows no significant direct effect at the 5% level on the preference parameters r and β for either gender, suggesting that IQ operates primarily through deliberation noise and guessing rather than through preferences.

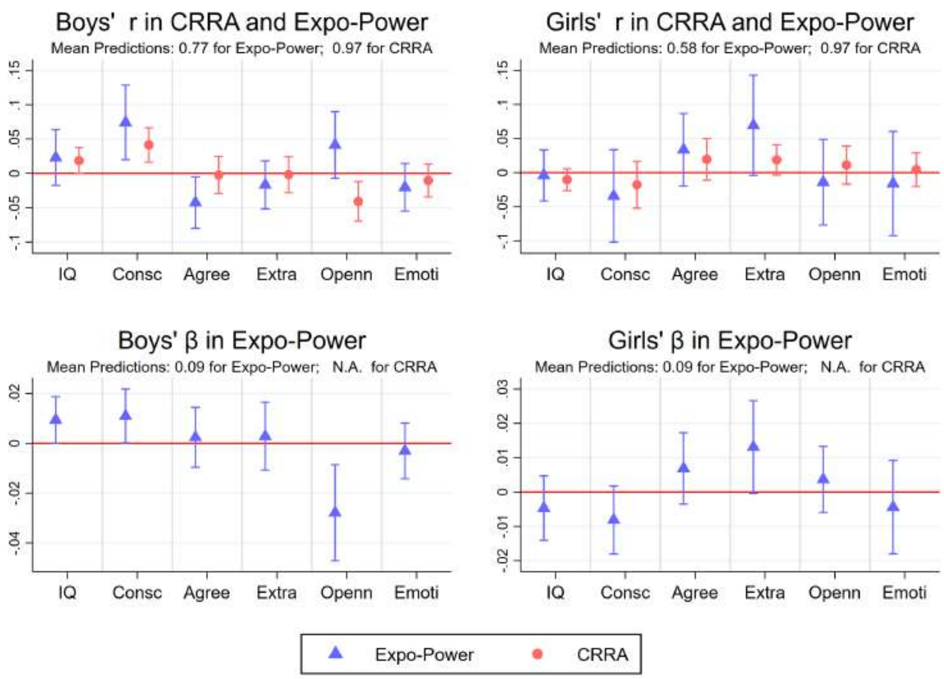


Figure P.1: Covariate Effects on Risk Aversion: Full Model Comparison

Notes: We compute predictions for each observation and average them over the sample. Coefficient estimates for effects of IQ and Big Five on $\text{RRA}(s+w) = r + \beta(s+w)^{1-r}$ evaluated at $w = 1$. Blue triangles: Expo-Power, red circles: CRRA. Boys: Conscientiousness increases risk aversion curvature r (coef. $= 0.074$, $p < 0.05$ under Expo-Power). Girls: no trait effect is statistically significant. Mean predictions, evaluated at $s + w = 1$ where $\text{RRA} = r + \beta$: boys 0.86 (Expo-Power) vs. 0.97 (CRRA), girls 0.67 (Expo-Power) vs. 0.97 (CRRA). Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

P.2 Deliberation Noise

Figure P.2 presents the effects of IQ and personality traits on transitory noise σ_{u^D} (top panels) and ICC^D (bottom panels). Lower noise and higher ICC indicate more precise and consistent decision-making.

Under Expo-Power, Conscientiousness reduces deliberation noise (-0.147, $p < 0.01$) for boys, indicating that conscientious boys make more precise decisions. IQ also reduces noise

(-0.099, $p < 0.10$), consistent with the interpretation that IQ reduces deliberation noise. Effects on ICC^D are weaker, with Emotional Stability showing a negative effect (-0.058, $p < 0.05$).

For girls, Extraversion reduces deliberation noise (-0.198, though not statistically significant at conventional levels), suggesting more extraverted girls make more precise decisions. IQ shows no significant effect on noise for girls, contrasting with boys.

The mean predictions highlight functional form sensitivity: boys show $\sigma_{u^D} = 0.58$ (Expo-Power) vs. 0.47 (CRRA), girls show $\sigma_{u^D} = 0.91$ (Expo-Power) vs. 0.52 (CRRA). The larger discrepancy for girls reflects their greater sensitivity to functional form restrictions.

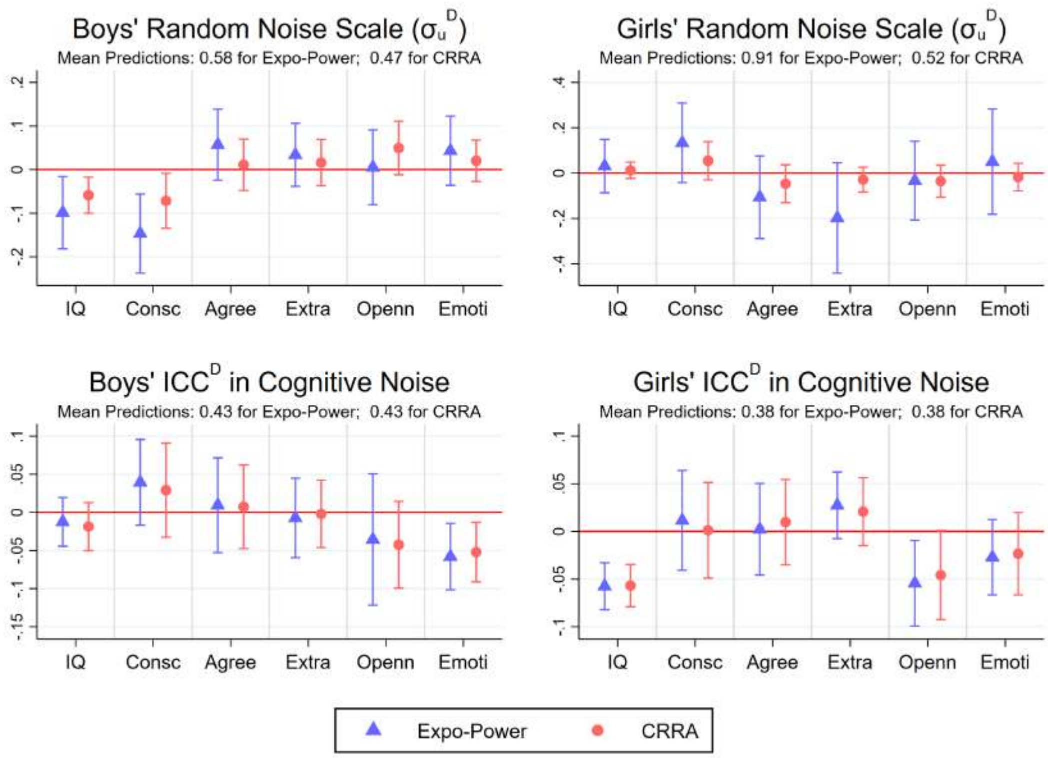


Figure P.2: Covariate Effects on Deliberation Noise: Noise and ICC

Notes: We compute predictions for each observation and average them over the sample. Top panels show effects on transitory noise σ_{u^D}, bottom panels show effects on ICC^D. Blue triangles: Expo-Power, red circles: CRRA. Mean predictions—Boys: $\sigma_{u^D} = 0.58$ (Expo-Power) vs. 0.47 (CRRA), ICC$^D = 0.43$ for both. Girls: $\sigma_{u^D} = 0.91$ (Expo-Power) vs. 0.52 (CRRA), ICC$^D = 0.38$ for both. Error bars show 90% confidence intervals.

P.3 Guessing Structure

Figure P.3 displays the effects of IQ and personality traits on the structure of guessing behavior, comparing the Expo-Power and CRRA specifications. The top panels report effects on the guessing intra-class correlation ICC^G, which measures the persistence of individual differences in guessing, while the bottom panels report effects on the correlation $\rho = \text{Corr}(\theta^D, \theta^G)$ between the persistent components of deliberation and guessing noise.

Boys display a negative baseline correlation ($\rho = -0.15$) and girls a positive one ($\rho = +0.11$), with IQ strengthening the positive correlation among girls (coef. $= 0.087$, $p < 0.01$). The corresponding discussion appears in the main text (Section 5.3).

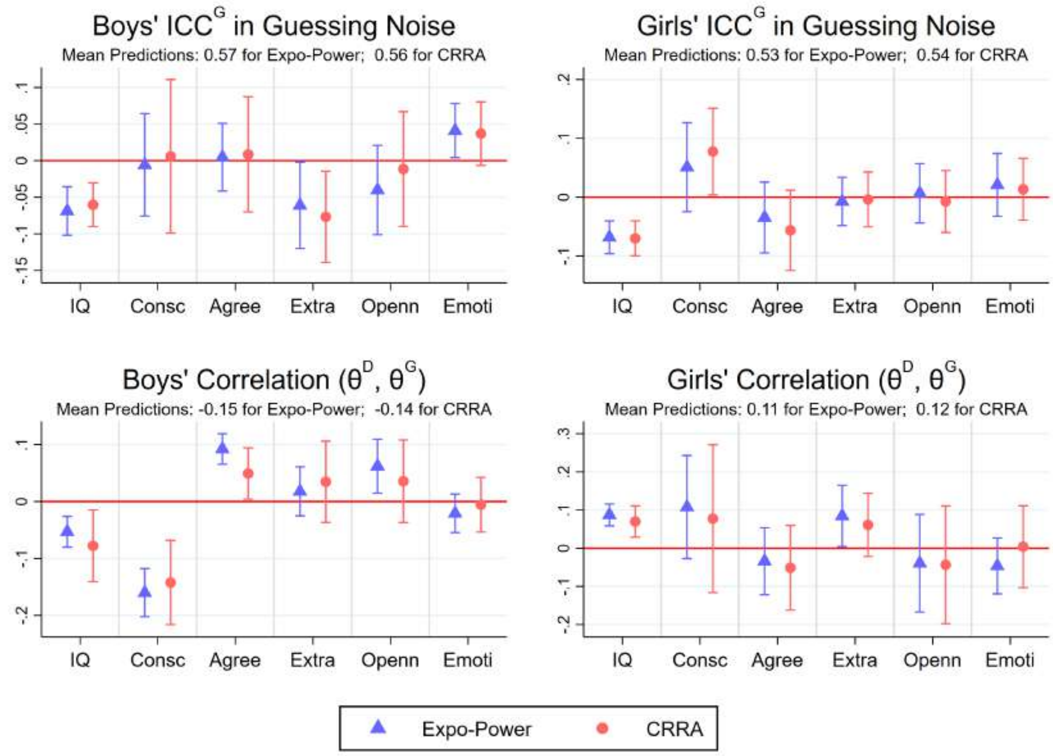


Figure P.3: Guessing Structure: ICC and Error Correlation

Notes: We compute predictions for each observation and average them over the sample. Top panels show effects on ICC^G, bottom panels show effects on correlation $\rho = \text{Corr}(\theta^D, \theta^G)$. Boys: Emotional Stability increases ICC^G (coef. $= 0.041$, $p < 0.10$), baseline correlation negative ($\rho = -0.15$). Girls: baseline correlation positive ($\rho = +0.11$), IQ strengthens it (coef. $= 0.087$, $p < 0.01$). Error bars show 90% confidence intervals.

P.4 Guessing Probability

Figure P.4 displays the effects of IQ and personality traits on the overall probability of guessing, comparing the Expo-Power and CRRA specifications. IQ strongly reduces guessing for both genders (boys: -0.069, girls: -0.096, both $p < 0.01$), while personality traits have weak effects on overall guessing frequency—only Openness significantly reduces guessing among boys (-0.040, $p < 0.05$). Mean guessing probabilities are 14% for boys and 19% for girls. The corresponding discussion appears in the main text (Section 5.3).

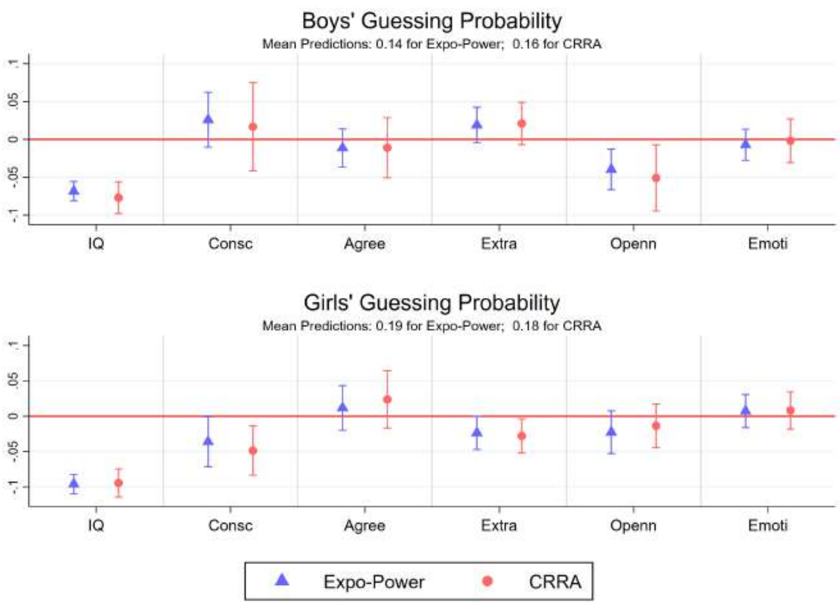


Figure P.4: Trait Effects on Guessing Probability

Notes: We compute predictions for each observation and average them over the sample. Effects of IQ and traits on overall guessing probability. Blue triangles: Expo-Power, red circles: CRRA. Boys: mean 14%, Girls: mean 19%. IQ strongly reduces guessing for both genders. Absence of strong trait effects on overall frequency contrasts with significant effects on ICC^G and correlation (Figure P.3), indicating traits affect consistency and structure rather than frequency. Error bars show 90% confidence intervals.

Q Alternative Guessing Specifications

This appendix presents detailed parameter estimates and coefficient plots comparing three alternative specifications of guessing behavior. Model 1 (baseline) allows personality traits and IQ to affect guessing propensity directly. Model 2 specifies that deliberation noise

(σ_{u^D}) affects guessing but traits do not. Model 3 activates both channels, allowing both σ_{u^D} and traits to affect guessing simultaneously. All models incorporate ICC structure and use teacher-reported personality measures.

We present results for the Expo-Power utility specification (Section Q.1). In the coefficient plots, blue circles represent Model 1, red squares represent Model 2, and green triangles represent Model 3.

Q.1 Expo-Power Specification

Q.1.1 Parameter Estimates: Expo-Power Models

Table Q.1: Boys, Expo-Power, Model 1: Parameter Estimates

	Parameter						
	$r(\boldsymbol{X})$ (1)	$\beta(\boldsymbol{X})$ (2)	$\sigma_{u^D}(\boldsymbol{X})$ (3)	$\text{ICC}^D(\boldsymbol{X})$ (4)	$\text{ICC}^G(\boldsymbol{X})$ (5)	$\text{Corr}(\theta^D,\theta^G)(\boldsymbol{X})$ (6)	$\text{Pr(Guess)}(\boldsymbol{X})$ (7)
Constant	0.771^{***}	0.091^{***}	0.577^{***}	0.428^{***}	0.574^{***}	-0.150^{***}	0.144^{***}
	(0.033)	(0.009)	(0.046)	(0.033)	(0.025)	(0.034)	(0.014)
IQ	0.023	0.009^{*}	-0.099^{*}	-0.012	-0.069^{***}	-0.053^{***}	-0.069^{***}
	(0.025)	(0.006)	(0.050)	(0.019)	(0.020)	(0.016)	(0.008)
Conscientiousness	0.074^{**}	0.011^{*}	-0.147^{***}	0.039	-0.006	-0.160^{***}	0.026
	(0.033)	(0.007)	(0.055)	(0.034)	(0.043)	(0.026)	(0.022)
Agreeableness	-0.042^{*}	0.002	0.057	0.009	0.005	0.092^{***}	-0.011
	(0.023)	(0.007)	(0.049)	(0.038)	(0.028)	(0.016)	(0.015)
Extraversion	-0.017	0.003	0.034	-0.007	-0.061^{*}	0.018	0.019
	(0.021)	(0.008)	(0.044)	(0.032)	(0.036)	(0.026)	(0.014)
Openness	0.041	-0.028^{**}	0.005	-0.036	-0.040	0.062^{**}	-0.040^{**}
	(0.030)	(0.012)	(0.052)	(0.052)	(0.037)	(0.029)	(0.016)
Emotional Stability	-0.020	-0.003	0.043	-0.058^{**}	0.041^{*}	-0.021	-0.007
	(0.021)	(0.007)	(0.048)	(0.027)	(0.023)	(0.021)	(0.013)

Notes: Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Model 1 uses the guessing specification in equation (21), where IQ and personality traits enter the guessing probability of equation (8) directly through $\eta(\boldsymbol{X}_i)$. Each parameter depends on covariates as in equations (18) and (19).

Table Q.2: Boys, Expo-Power, Model 2: Parameter Estimates

	Parameter							
	$r(\boldsymbol{X})$	$\beta(\boldsymbol{X})$	$\sigma_{u^D}(\boldsymbol{X})$	$\text{ICC}^D(\boldsymbol{X})$	$\text{ICC}^G(\boldsymbol{X})$	$\text{Corr}(\theta^D, \theta^G)(\boldsymbol{X})$	Pr(Guess)$(\boldsymbol{X})$	Effect of σ_{u^D} on Guessing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	0.772^{***}	0.089^{***}	0.584^{***}	0.420^{***}	0.597^{***}	-0.135^{***}	0.116^{***}	0.296
	(0.032)	(0.009)	(0.055)	(0.032)	(0.027)	(0.025)	(0.017)	(0.276)
IQ	0.024	0.007	-0.102	-0.017	-0.088^{***}	-0.094^{***}	-0.034^{***}	
	(0.022)	(0.008)	(0.066)	(0.027)	(0.026)	(0.014)	(0.011)	
Conscientiousness	0.054^{*}	0.009	-0.079	0.003	0.044^{*}	-0.175^{***}	-0.026^{***}	
	(0.029)	(0.008)	(0.066)	(0.034)	(0.025)	(0.024)	(0.010)	
Agreeableness	-0.036^{**}	0.005	0.028	0.031	-0.013	0.068^{***}	0.009	
	(0.018)	(0.007)	(0.037)	(0.031)	(0.026)	(0.014)	(0.009)	
Extraversion	-0.024	0.001	0.053	-0.018	-0.072^{**}	0.141^{***}	0.018	
	(0.017)	(0.007)	(0.033)	(0.026)	(0.036)	(0.021)	(0.016)	
Openness	0.060^{***}	-0.026^{**}	-0.043	-0.011	-0.049	-0.022	-0.014	
	(0.021)	(0.011)	(0.035)	(0.039)	(0.043)	(0.020)	(0.021)	
Emotional Stability	-0.015	-0.003	0.023	-0.046^{**}	0.012	0.025	0.008	
	(0.016)	(0.006)	(0.035)	(0.019)	(0.028)	(0.017)	(0.011)	

Notes: Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Model 2 uses the guessing specification in equation (22), where the deliberation noise σ_{u^D} of equation (19) enters the guessing probability of equation (8) directly and IQ and personality traits enter only indirectly through their effect on σ_{u^D}. Column (8) reports η_σ, the coefficient on σ_{u^D} in equation (22).

Table Q.3: Boys, Expo-Power, Model 3: Parameter Estimates

	Parameter							
	$r(\boldsymbol{X})$	$\beta(\boldsymbol{X})$	$\sigma_{u^D}(\boldsymbol{X})$	$\text{ICC}^D(\boldsymbol{X})$	$\text{ICC}^G(\boldsymbol{X})$	$\text{Corr}(\theta^D, \theta^G)(\boldsymbol{X})$	Pr(Guess)$(\boldsymbol{X})$	Effect of σ_{u^D} on Guessing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	0.771^{***}	0.092^{***}	0.572^{***}	0.431^{***}	0.570^{***}	-0.146^{***}	0.147^{***}	-0.410
	(0.034)	(0.010)	(0.050)	(0.035)	(0.027)	(0.041)	(0.018)	(0.983)
IQ	0.032	0.010	-0.110^{*}	-0.013	-0.058	-0.060^{**}	-0.074^{***}	
	(0.031)	(0.007)	(0.063)	(0.021)	(0.055)	(0.024)	(0.019)	
Conscientiousness	0.063	0.009	-0.115	0.028	0.011	-0.173^{**}	0.017	
	(0.047)	(0.008)	(0.094)	(0.042)	(0.075)	(0.073)	(0.037)	
Agreeableness	-0.039	0.003	0.044	0.014	-0.005	0.103^{**}	-0.004	
	(0.024)	(0.008)	(0.050)	(0.037)	(0.035)	(0.044)	(0.024)	
Extraversion	-0.014	0.004	0.023	-0.004	-0.068	0.020	0.021	
	(0.025)	(0.008)	(0.050)	(0.033)	(0.046)	(0.038)	(0.016)	
Openness	0.032	-0.032^{**}	0.036	-0.046	-0.041	0.058	-0.044^{**}	
	(0.044)	(0.015)	(0.101)	(0.060)	(0.039)	(0.038)	(0.019)	
Emotional Stability	-0.010	0.001	0.013	-0.048	0.042^{*}	-0.018	-0.003	
	(0.044)	(0.014)	(0.112)	(0.043)	(0.025)	(0.043)	(0.019)	

Notes: Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Model 3 uses the guessing specification in equation (23), which combines the direct trait channel of equation (21) with the deliberation noise channel of equation (22). Column (8) reports η_σ, the coefficient on σ_{u^D}.

Table Q.4: Girls, Expo-Power, Model 1: Parameter Estimates

	Parameter						
	$r(\boldsymbol{X})$ (1)	$\beta(\boldsymbol{X})$ (2)	$\sigma_{u^D}(\boldsymbol{X})$ (3)	$\text{ICC}^D(\boldsymbol{X})$ (4)	$\text{ICC}^G(\boldsymbol{X})$ (5)	$\text{Corr}(\theta^D, \theta^G)(\boldsymbol{X})$ (6)	$\Pr(\text{Guess})(\boldsymbol{X})$ (7)
Constant	0.576***	0.093***	0.908***	0.383***	0.531***	0.108**	0.187***
	(0.025)	(0.005)	(0.082)	(0.019)	(0.023)	(0.045)	(0.012)
IQ	−0.004	−0.005	0.031	−0.057***	−0.068***	0.087***	−0.096***
	(0.023)	(0.006)	(0.072)	(0.015)	(0.017)	(0.018)	(0.008)
Conscientiousness	−0.034	−0.008	0.134	0.012	0.051	0.108	−0.036*
	(0.041)	(0.006)	(0.107)	(0.032)	(0.046)	(0.082)	(0.021)
Agreeableness	0.034	0.007	−0.107	0.002	−0.034	−0.034	0.012
	(0.032)	(0.006)	(0.111)	(0.029)	(0.037)	(0.053)	(0.019)
Extraversion	0.070	0.013	−0.198	0.027	−0.007	0.085*	−0.024*
	(0.045)	(0.008)	(0.148)	(0.021)	(0.025)	(0.049)	(0.014)
Openness	−0.014	0.004	−0.034	−0.054**	0.007	−0.039	−0.023
	(0.038)	(0.006)	(0.106)	(0.027)	(0.031)	(0.078)	(0.018)
Emotional Stability	−0.016	−0.004	0.051	−0.027	0.021	−0.046	0.007
	(0.047)	(0.008)	(0.141)	(0.024)	(0.032)	(0.044)	(0.014)

Notes: Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Model 1 uses the guessing specification in equation (21), where IQ and personality traits enter the guessing probability of equation (8) directly through $\eta(\boldsymbol{X}_i)$. Each parameter depends on covariates as in equations (18) and (19).

Table Q.5: Girls, Expo-Power, Model 2: Parameter Estimates

	Parameter							
	$r(\boldsymbol{X})$ (1)	$\beta(\boldsymbol{X})$ (2)	$\sigma_{u^D}(\boldsymbol{X})$ (3)	$\text{ICC}^D(\boldsymbol{X})$ (4)	$\text{ICC}^G(\boldsymbol{X})$ (5)	$\text{Corr}(\theta^D, \theta^G)(\boldsymbol{X})$ (6)	$\Pr(\text{Guess})(\boldsymbol{X})$ (7)	Effect of σ_{u^D} on Guessing (8)
Constant	0.578***	0.093***	0.875***	0.385***	0.530***	0.121***	0.185***	2.857*
	(0.023)	(0.004)	(0.064)	(0.019)	(0.023)	(0.046)	(0.013)	(1.622)
IQ	0.009	0.000	−0.032*	−0.052***	−0.070***	0.078***	−0.090***	
	(0.012)	(0.002)	(0.017)	(0.014)	(0.020)	(0.020)	(0.011)	
Conscientiousness	0.003	0.002	−0.013	0.023	0.063	0.104	−0.038*	
	(0.027)	(0.004)	(0.010)	(0.028)	(0.047)	(0.083)	(0.023)	
Agreeableness	0.005	−0.001	0.005	−0.006	−0.038	−0.064	0.013	
	(0.022)	(0.003)	(0.008)	(0.024)	(0.041)	(0.064)	(0.023)	
Extraversion	0.017	0.002	−0.011	0.015	−0.009	0.077**	−0.031**	
	(0.018)	(0.002)	(0.009)	(0.021)	(0.027)	(0.036)	(0.013)	
Openness	−0.019	0.001	−0.005	−0.049*	−0.001	−0.065	−0.015	
	(0.024)	(0.003)	(0.006)	(0.027)	(0.032)	(0.061)	(0.017)	
Emotional Stability	−0.005	0.000	−0.000	−0.029	0.024	−0.009	−0.000	
	(0.020)	(0.003)	(0.005)	(0.025)	(0.030)	(0.047)	(0.015)	

Notes: Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Model 2 uses the guessing specification in equation (22), where the deliberation noise σ_{u^D} of equation (19) enters the guessing probability of equation (8) directly and IQ and personality traits enter only indirectly through their effect on σ_{u^D}. Column (8) reports η_σ, the coefficient on σ_{u^D} in equation (22).

Table Q.6: Girls, Expo-Power, Model 3: Parameter Estimates

	Parameter							
	$r(\boldsymbol{X})$	$\beta(\boldsymbol{X})$	$\sigma_{u^D}(\boldsymbol{X})$	$\text{ICC}^D(\boldsymbol{X})$	$\text{ICC}^G(\boldsymbol{X})$	$\text{Corr}(\theta^D, \theta^G)(\boldsymbol{X})$	$\text{Pr(Guess)}(\boldsymbol{X})$	Effect of σ_{u^D} on Guessing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	0.578^{***}	0.094^{***}	0.902^{***}	0.382^{***}	0.529^{***}	0.109^{*}	0.185^{***}	-0.658
	(0.024)	(0.005)	(0.080)	(0.019)	(0.025)	(0.062)	(0.014)	(1.051)
IQ	-0.001	-0.004	0.020	-0.057^{***}	-0.066^{***}	0.081^{***}	-0.096^{***}	
	(0.031)	(0.008)	(0.104)	(0.016)	(0.021)	(0.021)	(0.011)	
Conscientiousness	-0.031	-0.007	0.117	0.013	0.049	0.123	-0.038	
	(0.034)	(0.006)	(0.081)	(0.030)	(0.058)	(0.118)	(0.025)	
Agreeableness	0.015	0.003	-0.046	-0.001	-0.048	-0.050	0.022	
	(0.031)	(0.006)	(0.090)	(0.026)	(0.048)	(0.067)	(0.025)	
Extraversion	0.070^{**}	0.013^{***}	-0.206^{**}	0.031	-0.014	0.086^{*}	-0.024	
	(0.033)	(0.005)	(0.100)	(0.021)	(0.034)	(0.050)	(0.017)	
Openness	-0.017	0.003	-0.020	-0.055^{**}	0.014	-0.070	-0.023	
	(0.038)	(0.007)	(0.118)	(0.027)	(0.036)	(0.060)	(0.018)	
Emotional Stability	-0.014	-0.004	0.049	-0.028	0.027	-0.032	0.004	
	(0.034)	(0.006)	(0.099)	(0.024)	(0.044)	(0.083)	(0.016)	

Notes: Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Model 3 uses the guessing specification in equation (23), which combines the direct trait channel of equation (21) with the deliberation noise channel of equation (22). Column (8) reports η_σ, the coefficient on σ_{u^D}.

Q.1.2 Coefficient Plots: Expo-Power Models

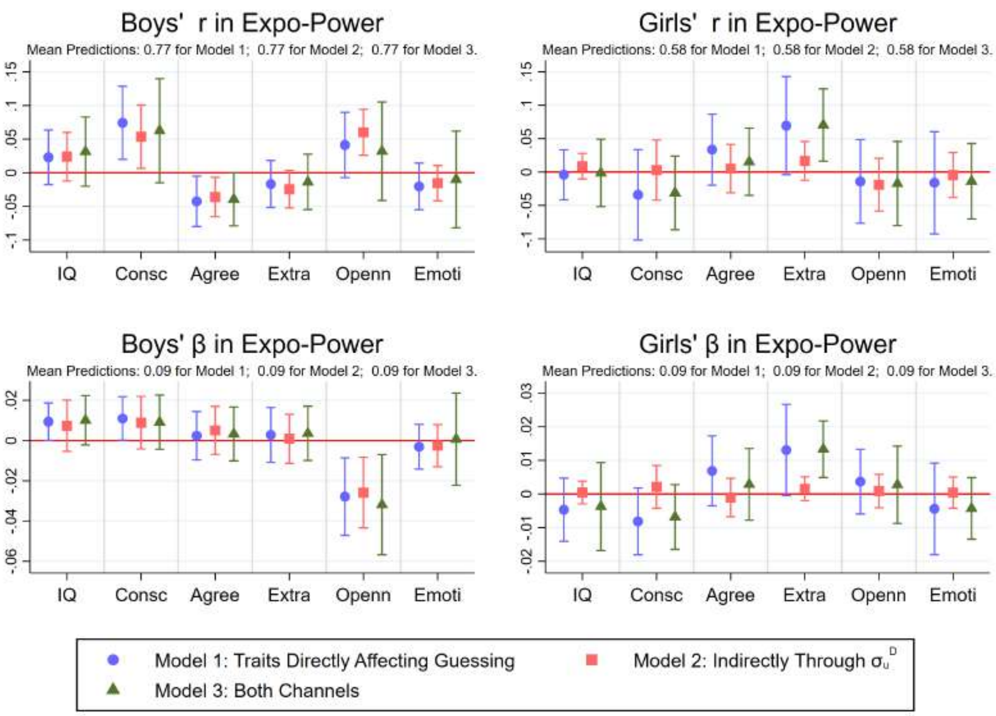


Figure Q.1: Expo-Power Parameters r and β Across Guessing Models

Notes: We compute predictions for each observation and average them over the sample. Blue circles: Model 1 (traits affect guessing). Red squares: Model 2 (σ_{u^D} affects guessing). Green triangles: Model 3 (both channels). Boys and girls are shown. Mean predictions across the three models appear in each panel's subtitle. Error bars show 90% confidence intervals.

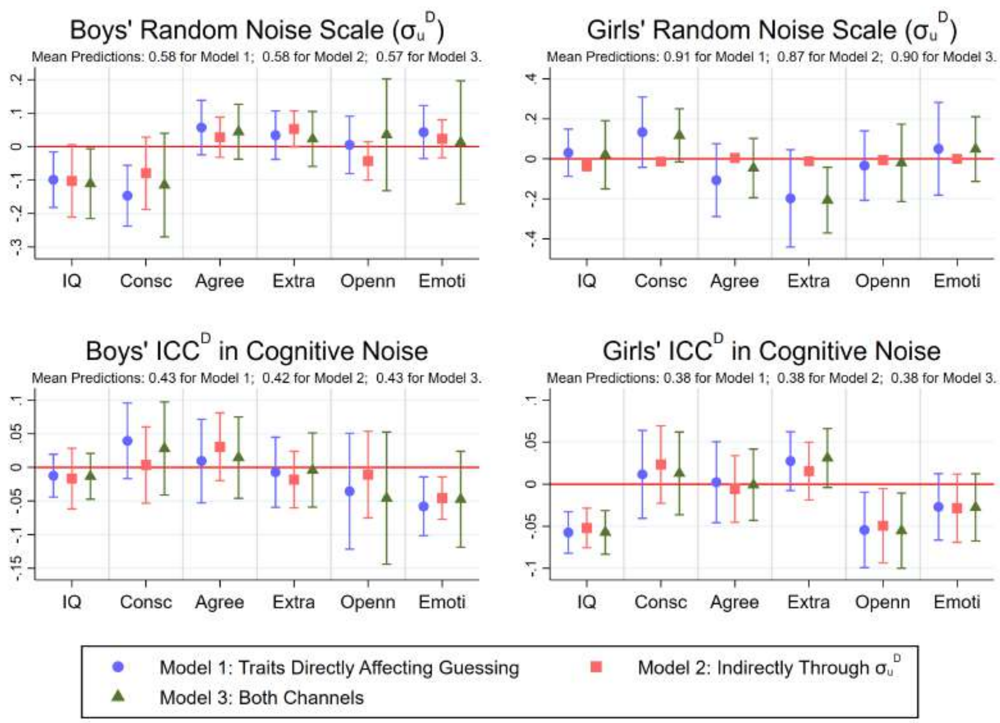


Figure Q.2: Transitory Noise σ_{u^D} and ICC^D Across Guessing Models

Notes: We compute predictions for each observation and average them over the sample. Blue circles: Model 1 (traits affect guessing). Red squares: Model 2 (σ_{u^D} affects guessing). Green triangles: Model 3 (both channels). Boys and girls are shown. Mean predictions across the three models appear in each panel's subtitle. Error bars show 90% confidence intervals.

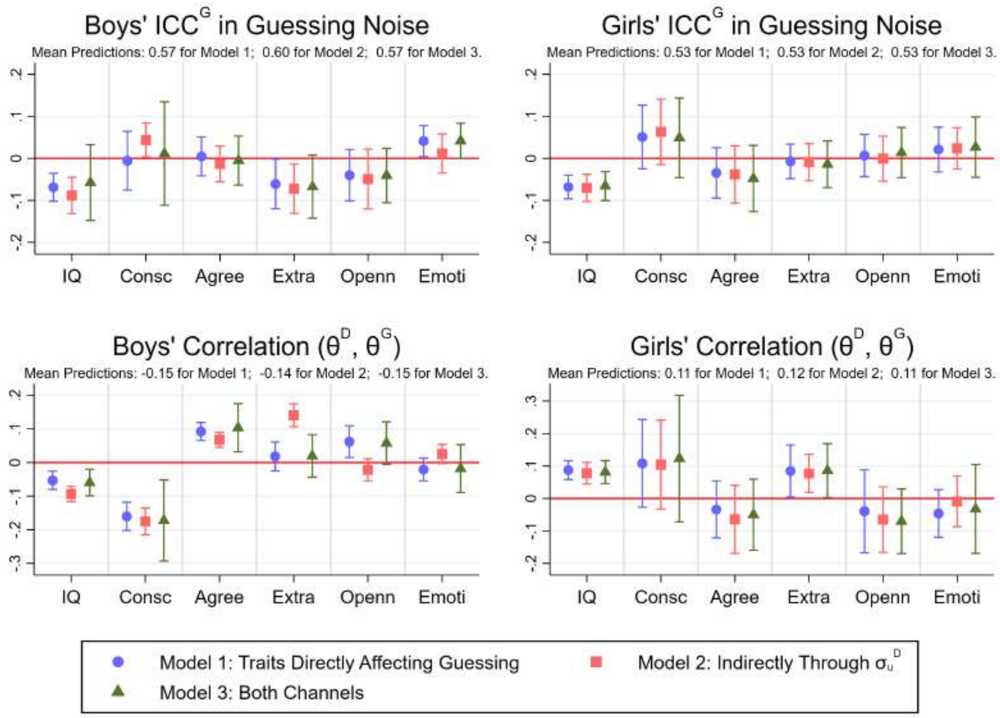


Figure Q.3: ICC^G and Deliberation–Guessing Correlation Across Guessing Models

Notes: We compute predictions for each observation and average them over the sample. Blue circles: Model 1 (traits affect guessing). Red squares: Model 2 (σ_{u^D} affects guessing). Green triangles: Model 3 (both channels). Boys and girls are shown. Mean predictions across the three models appear in each panel's subtitle. Error bars show 90% confidence intervals.

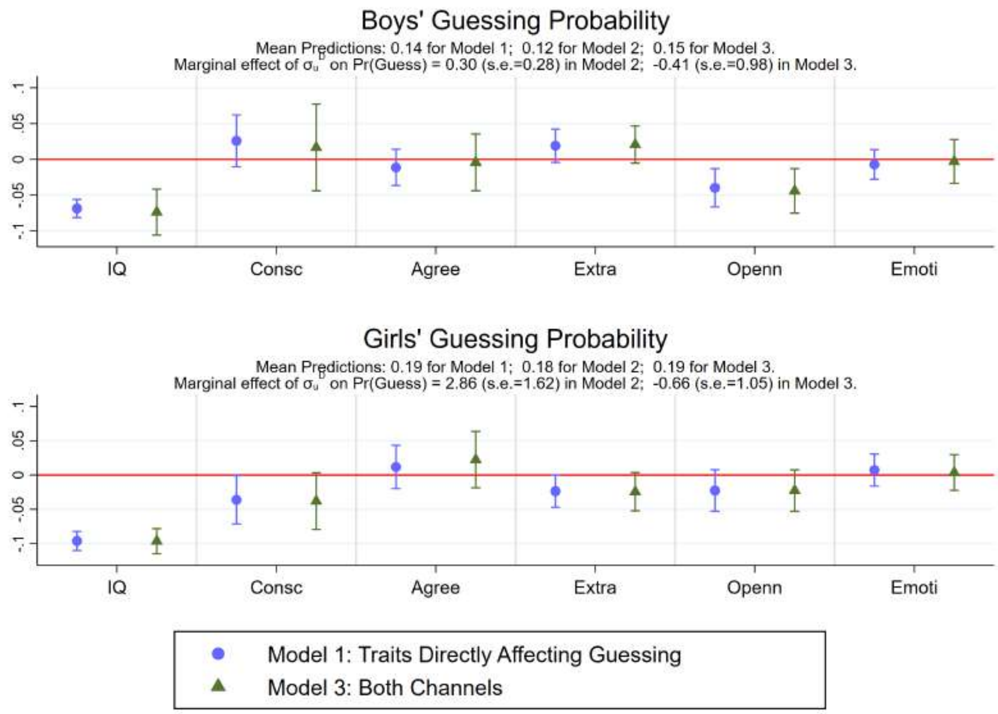


Figure Q.4: Guessing Probability Across Guessing Models

Notes: We compute predictions for each observation and average them over the sample. Blue circles: Model 1 (traits affect guessing). Red squares: Model 2 (σ_{u^D} affects guessing). Green triangles: Model 3 (both channels). Boys and girls are shown. Mean predictions across the three models appear in each panel's subtitle. Error bars show 90% confidence intervals.

R Component Interactions

This appendix documents the interactions between model components. Because interaction between non-additive components is the norm rather than the exception in nonlinear models, we present these results as technical documentation rather than as a substantive finding.

Testing whether the various components' contributions are additive reveals substantial positive interaction. For boys' Expo-Power, components estimated separately yield 2,375.8

points (IQ + Personality: 382.4, ICC: 1,963.6, Guessing: 28.7, β parameter: 1.0), yet the full model improvement is 3,261.1 points. The 885.3-point difference represents 37.3% more explanatory power than simple additive contributions. For girls, the component interaction is 618.4 points (28.6% beyond the 2,161.3-point sum), as shown in Figure R.1.

The β parameter contribution reveals a gender asymmetry: adding wealth-dependent risk aversion improves fit by merely 1.0 log-likelihood point for boys but 77.3 points for girls—a 77-fold difference. This is why CRRA's constant-curvature restriction binds girls' preferences far more tightly than boys': girls rely on the wealth-dependence that CRRA rules out, while boys' preferences are left essentially unconstrained. The reversal in the estimated gender gap itself, however, operates primarily through the risk aversion curvature r, not through β directly, as discussed below.

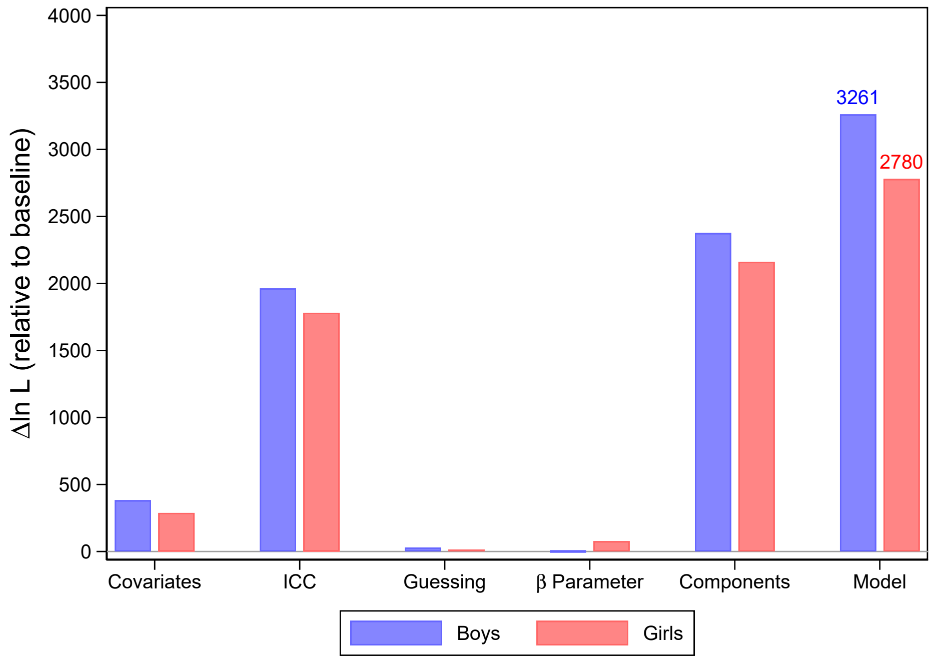


Figure R.1: Component Contributions to Log-Likelihood Improvements

Notes: Log-likelihood improvements relative to baseline CRRA for boys (blue) and girls (red). The first four pairs of bars show improvements when each component is added separately. The fifth pair of bars, "Components," shows the sum of the first four pairs. The sixth pair of bars, "Model," shows the actual improvement achieved when all components are included jointly. Component interactions—the gap between the sum of the individual contributions and their joint contribution—demonstrate that the components interact positively: 885 points (37.3%) for boys, 618 points (28.6%) for girls. Boys: N = 1,115, Girls: N = 1,089.

The positive component interactions indicate that model components interact synergistically rather than independently. This pattern is expected in nonlinear models for several reasons. First, ICC structure changes the interpretation of other parameters by separating persistent from transitory variance. Second, guessing behavior interacts with ICC through the correlation between deliberation and guessing noise. Third, the β parameter's contribution depends on how well other components have isolated true preference heterogeneity from deliberation noise.

The 77-fold difference in β parameter contribution between boys (1.0 points) and girls

(77.3 points) reflects the greater importance of wealth-dependent risk aversion for girls. This difference in the β contribution does not directly explain the gender gap reversal, which operates primarily through the risk aversion curvature r. Rather, it is what makes CRRA's constant-curvature restriction bind girls' preferences far more tightly than boys'.

The component-interaction finding is robust to alternative orderings of component addition, to using CRRA instead of Expo-Power as the functional form, and to excluding individual personality traits. These results confirm that accurate preference recovery requires simultaneous modeling of preferences, deliberation noise, and guessing—partial models that include only some components will systematically misestimate all parameters.

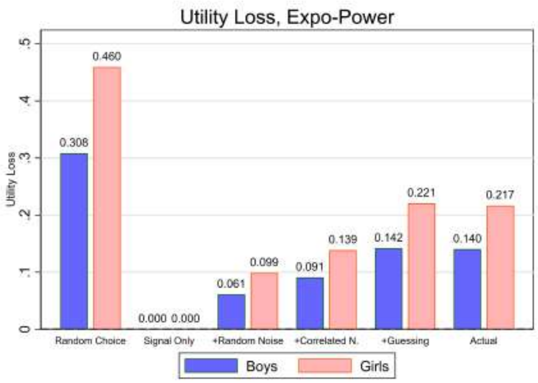
S Decomposition of Choice Mistakes and Utility Loss

Figure S.1 presents the sequential decomposition of choice mistakes under the Expo-Power specification.[30] Each panel displays benchmarks representing progressive inclusion of model components, in the following order: (1) the level under purely random choice (a mistake rate of 0.50 if children flipped coins), (2) the level implied by true preferences alone (zero mistakes), (3) the simulated level with transitory random deliberation noise (u_{ij}^D) only, (4) the level adding persistent individual noise (θ_i^D), (5) the level with all three components including guessing (omitted in the no-guessing panel), and (6) the level in observed choices.

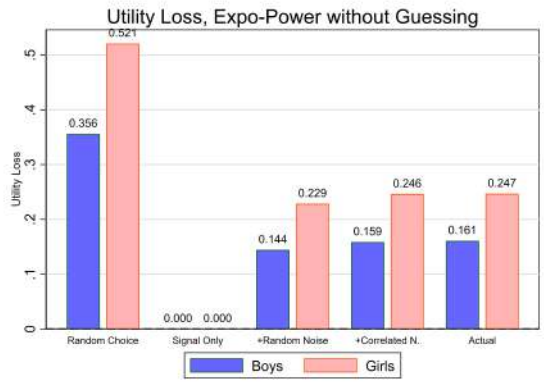
Both boys and girls exhibit mistake rates around 30%, substantially below the 50% randomization benchmark but far above zero. Random transitory noise contributes approximately 19–21% to the total mistake rate, persistent individual noise adds 4–5%, and guessing contributes approximately 6%. These sequential shares are order-dependent and differ from the Shapley shares reported in the main text (where guessing accounts for roughly 9%), since the sequential decomposition credits each component only with its marginal contribution given those already included. In utility-loss terms, the Shapley decomposition attributes the largest single contribution to guessing (0.062 for boys and 0.100

[30] Results for CRRA are very similar (see Figure S.2).

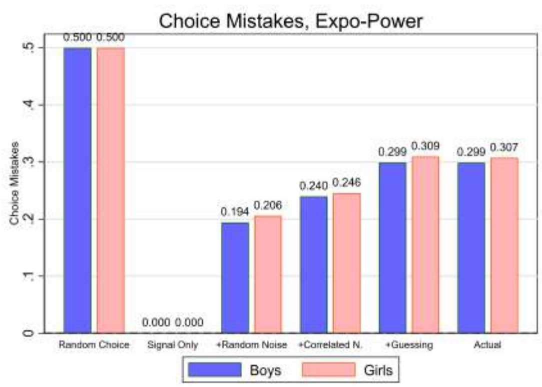
for girls), with total utility loss of 0.142 and 0.221 respectively under Expo-Power. The close alignment between full model predictions and actual observed choices validates the model's ability to capture decision error structure.



(a) Utility Loss



(b) Utility Loss in a Model without Guessing

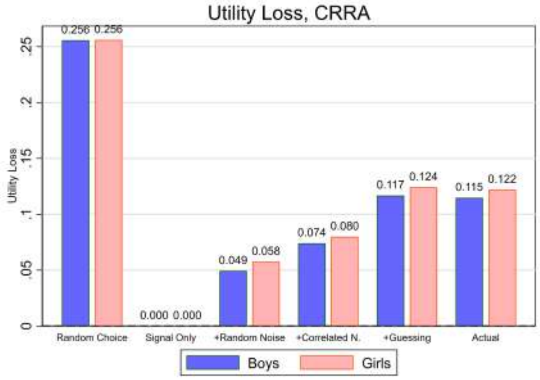


(c) Choice Mistakes

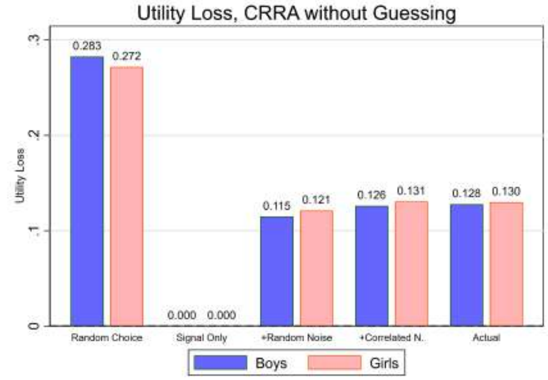
Figure S.1: Sequential Decomposition of Choice Mistakes and Utility Loss (Expo-Power)

Notes: Sequential decomposition for utility loss (panel a), for utility loss in a model estimated without guessing (panel b), and for choice mistakes (panel c). Blue bars represent boys, red bars represent girls. "Random Choice" is the benchmark when children choose by coin flip. "Signal Only" is the value under perfect decision-making. "+ Random Noise" introduces transitory noise (u_{ij}^D), "+ Persistent N." adds persistent noise (θ_i^D), and "+ Guessing" activates the remaining guessing components (u_{ij}^G, θ_i^G, and the correlation $\rho(\theta_i^D, \theta_i^G)$). "Actual" is the empirical value. In panel (b), with no guessing component, the transitory and persistent noise terms alone are estimated to absorb the full utility loss, illustrating how omitting guessing reattributes its contribution to noise.

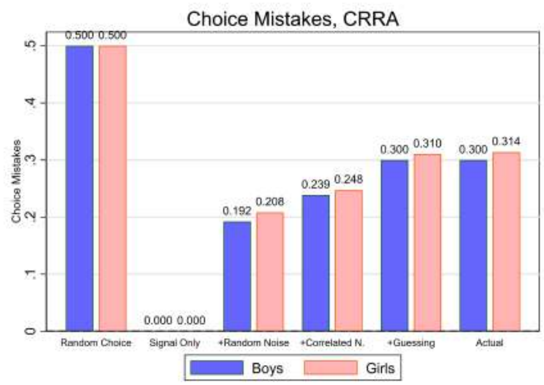
The CRRA versions of these decompositions are very similar to the Expo-Power specifications: the sequential decomposition (Figure S.2) and the Shapley decomposition (Figure S.3). Random transitory noise is the largest single source of mistakes, persistent individual noise and guessing each contribute a substantial share, and total mistake rates are around 30% for both genders. The two specifications diverge for *utility loss*: under CRRA the Shapley decomposition attributes 0.051 (boys) and 0.054 (girls) to guessing, with totals of 0.117 and 0.124, below the corresponding Expo-Power values of 0.062 and 0.100 for guessing and 0.142 and 0.221 in total.



(a) Utility Loss



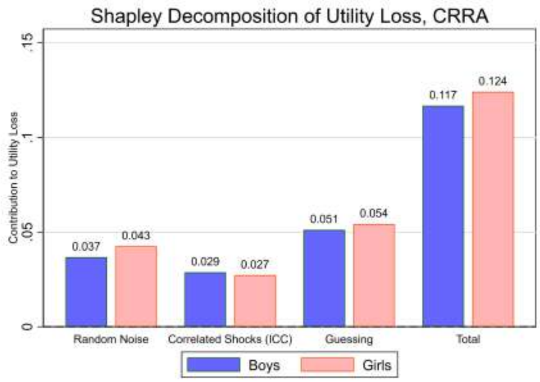
(b) Utility Loss in a Model without Guessing



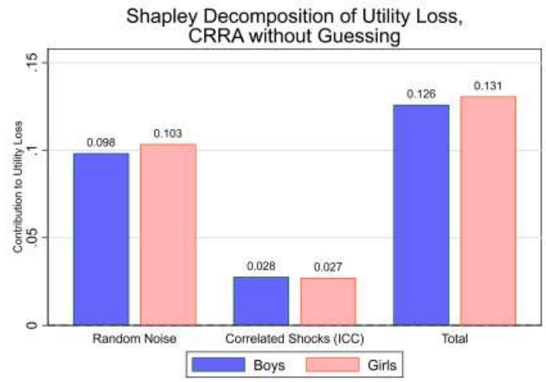
(c) Choice Mistakes

Figure S.2: Sequential Decomposition of Choice Mistakes and Utility Loss (CRRA)

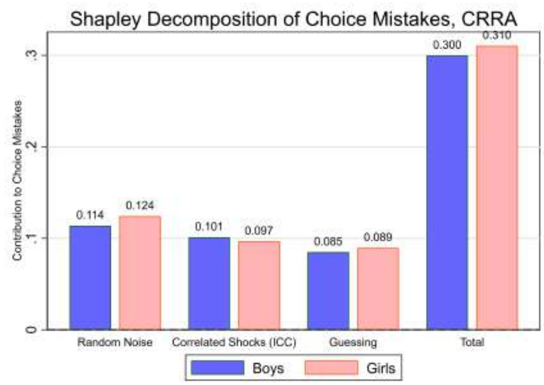
Notes: Sequential decomposition for utility loss (panel a), for utility loss in a model estimated without guessing (panel b), and for choice mistakes (panel c). Blue bars represent boys, red bars represent girls. "Random Choice" is the benchmark when children choose by coin flip. "Signal Only" is the value under perfect decision-making. "+ Random Noise" introduces transitory noise (u_{ij}^D), "+ Persistent N." adds persistent noise (θ_i^D), and "+ Guessing" activates the remaining guessing components (u_{ij}^G, θ_i^G, and the correlation $\rho(\theta_i^D, \theta_i^G)$). "Actual" is the empirical value. In panel (b), with no guessing component, the transitory and persistent noise terms alone are estimated to absorb the full utility loss, illustrating how omitting guessing reattributes its contribution to noise.



(a) Utility Loss



(b) Utility Loss in a Model without Guessing



(c) Choice Mistakes

Figure S.3: Shapley Decomposition of Choice Mistakes and Utility Loss (CRRA)

Notes: Shapley value decomposition for utility loss (panel a), for utility loss in a model estimated without guessing (panel b), and for choice mistakes (panel c). Blue bars: boys, red bars: girls. Each component's contribution is its average marginal effect across all possible orderings. "Random Noise" refers to transitory noise u_{ij}^D, "Persistent Noise" to individual noise θ_i^D, and "Guessing" to guessing behavior. "Total" shows the sum. In panel (b), the model omits guessing, and its contribution is absorbed by the noise components.

T OLS Analysis of Inconsistencies of Choices

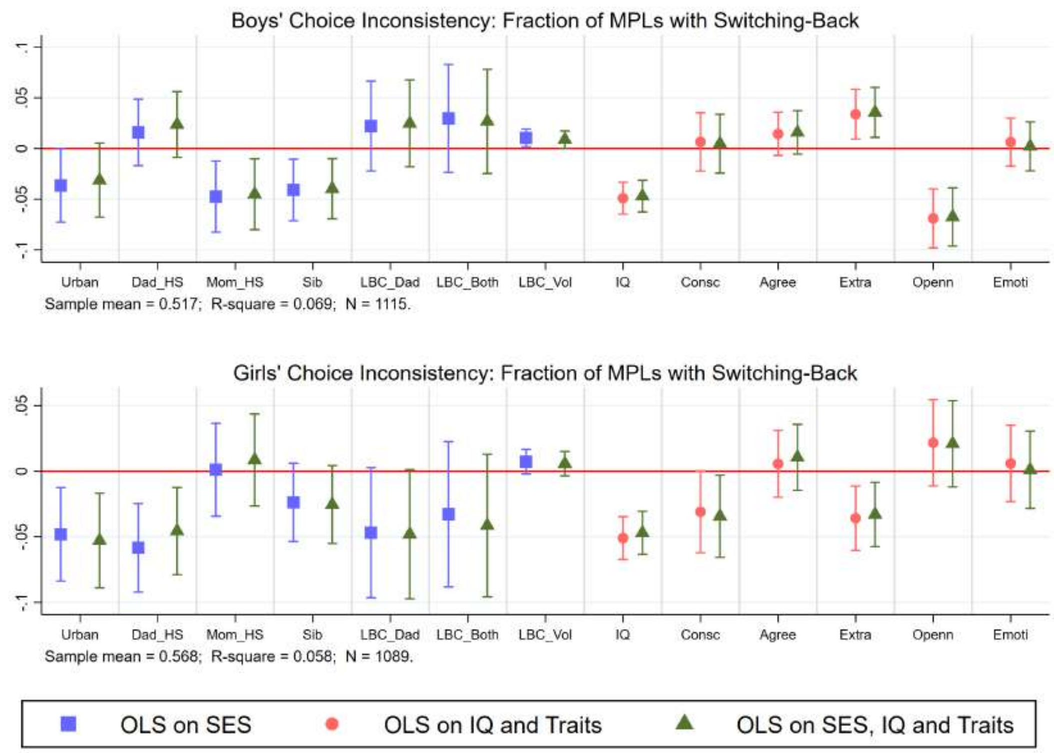


Figure T.1: Reduced-Form Analysis: Choice Inconsistency (Inconsistent Behavior)

Notes: This figure presents coefficient estimates from linear probability models where the dependent variable is the fraction of multiple price lists (MPLs) in which the child switched back to a lower-variance option after choosing a higher-variance one. Inconsistent behavior may indicate confusion, inattention, or inconsistent preferences. The top panel shows results for boys ($N = 1{,}115$), the bottom panel shows results for girls ($N = 1{,}089$). Blue squares represent regressions on socioeconomic status (SES) variables only. Red circles represent regressions on IQ and personality traits only. Green triangles represent regressions including all covariates. Error bars show 90% confidence intervals. These results correspond to Table E.8 in Appendix E.

Appendix References

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

Bartlett, Maurice S. 1937. "The statistical conception of mental factors." *British Journal of Psychology* 28 (1): 97–104.

Carneiro, Pedro, Karsten T. Hansen, and James J. Heckman. 2003. "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice." *International Economic Review* 44 (2): 361–422.

Chávez, Cesar, Shuaizhang Feng, James J. Heckman, and Zhe Yang. 2026. "Alternative Models for the Risk Preferences of Young Children." Department of Economics, University of Chicago.

Chernoff, Herman. 1954. "On the Distribution of the Likelihood Ratio." *The Annals of Mathematical Statistics* 25 (3): 573–578. https://doi.org/10.1214/aoms/1177728725.

Domencich, Thomas A., and Daniel L. McFadden. 1975. *Urban Travel Demand: A Behavioral Analysis.* 215. Contributions to Economic Analysis. Amsterdam: North-Holland Publishing Company. ISBN: 9780444108302.

Feng, Shuaizhang, Yujie Han, James J. Heckman, and Tim Kautz. 2022. "Comparing the reliability and predictive power of child, teacher, and guardian reports of noncognitive skills." *Proceedings of the National Academy of Sciences* 119 (6): e2113992119. https://doi.org/10.1073/pnas.2113992119.

Feng, Shuaizhang, Jun Hyung Kim, and Zhe Yang. 2026. "Effects of Childe0hood Peers on Personality Traits." Submitted July 1, 2022; accepted August 8, 2024; electronically published February 10, 2025, *Journal of Labor Economics* 44 (2): 413–445. https://doi.org/10.1086/732770.

Heckman, James J., Rodrigo Pinto, and Peter A. Savelyev. 2013. "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review* 103, no. 6 (October): 2052–2086. https://doi.org/10.1257/aer.103.6.2052.

Heckman, James J., and Edward J. Vytlacil. 2007. "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation." Chap. 70 in *Handbook of Econometrics,* edited by James J. Heckman and Edward E. Leamer, vol. 6B, 4779–4874. Amsterdam: Elsevier B. V. https://doi.org/10.1016/S1573-4412(07)06070-9.

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

Rothenberg, Thomas J. 1971. "Identification in Parametric Models." *Econometrica* 39 (3): 577–591.

Saha, Atanu. 1993. "Expo-Power Utility: A 'Flexible' Form for Absolute and Relative Risk Aversion." *American Journal of Agricultural Economics* 75 (4): 905–913. https://doi.org/10.2307/1243978.

Self, Steven G., and Kung-Yee Liang. 1987. "Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions." *Journal of the American Statistical Association* 82 (398): 605–610. https://doi.org/10.1080/01621459.1987.10478472.

Soto, Christopher J., and Oliver 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): 117–143. https://doi.org/10.1037/pspp0000096.

Thurstone, Louis L. 1927. "A law of comparative judgment." *Psychological Review* 34 (4): 273–286. https://doi.org/10.1037/h0070288.

Xie, Danyang. 2000. "Power Risk Aversion Utility Functions." Working Paper version: CEMA Working Paper No. 22 (Oct 1999), *Annals of Economics and Finance* 1, no. 2 (November): 265–282. https://aeconf.com/Articles/Nov2000/aef010202.pdf.