Relating Cognitive Skills and Personality Traits to Economic Preferences: A Study of Chinese Children[*]

Cesar Chavez[*] Shuaizhang Feng[†] James J. Heckman[*]
Zhe Yang[‡]

[*]Center for Economics of Human Development, University of Chicago
[†]Institute for Economic and Social Research, Jinan University.
[‡]China Economic Research Institute, Liaoning University

December 13, 2025

Abstract

We examine the relationship of cognitive skills and personality traits to children's risk preferences and to the precision with which those preferences are elicited using rich data on Chinese schoolchildren. 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 decision precision and guessing behavior. Variances in shocks to choice reflect persistent individual-specific noise rather than transitory shocks. Socioeconomic background has little direct influence on preferences or decision quality once those traits are accounted for. Young boys are considerably more risk-averse than girls.

[*]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 or NSF. We thank Glenn Harrison, Patrick Kyllonen, Tamara McGavock, John Rust and Michael Woodford for their comments on various drafts of the paper. We thank Ariyan Mishra and Arjun Pathy for providing excellent research assistance on this project. Online Appendix is available here.

JEL Codes: D12, D90, J24, I21
Keywords: Risk preferences; cognitive skills; personality traits; decision noise; children; structural estimation

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 component of human achievement and now recognizes that personality traits are also predictive.[1]

This paper studies whether, to what degree, and through which channels cognitive skills and personality traits shape the measured economic preferences of children. Both cognitive skills and personality traits predict measures of preference as well as the consistency and quality of decision-making in laboratory and field settings (Burks et al., 2009). The recent literature examines the measurement of preferences and the challenge of extracting true preferences when individuals face cognitive limitations or 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, while simpler elicitation formats can reduce noise at the cost of precision. Together, this literature suggests the importance of accounting for cognitive noise, personality

[1] See Almlund et al. (2011) and Heckman et al. (2025) 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).

traits, and elicitation decisions in understanding choice behavior.

Early work by Borghans et al. (2008) raised the question of how personality traits relate to economic preferences. Dohmen et al. (2011) collect survey data on 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 cognitive ability and risk aversion partly reflects decision noise rather than true preference differences. Jagelka (2024) links personality traits and preferences in a random preference model due to Thurstone (1927) that incorporates measurement error and cognitive imprecision. This approach yields stronger and more interpretable relationships than those found in earlier survey-based studies. Cross-agent experiments likewise show that part of the heterogeneity in observed choices is driven by decision noise (von Gaudecker et al., 2011). This evidence illustrates the potential value of models that distinguish preference heterogeneity from noise.

This paper analyzes uniquely rich data on Chinese schoolchildren. We elicit detailed measures of preferences, cognitive skills, personality traits, and family background. We have data on children from fourth grade into college, enabling the study of preference formation before labor-market experience or higher education come into play. In this paper, we focus on the choices under uncertainty of fifth graders. This is the first step of a longitudinal analysis that we plan to pursue. We develop a framework that we will apply in future work. Our paper contributes to an emerging literature on children's economic preferences. 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 and time preferences (Sutter et al., 2019), but the mechanisms underlying this heterogeneity are not systematically explained.

We show that part of the variation in preferences reflects differences in the precision rather than differences in true underlying preferences. This is consistent with previous work (Andersson et al., 2016, 2020; Jagelka, 2024). That research shows that individuals with higher cognitive ability tend to exhibit less decision noise and make more consistent choices (Andersson et al., 2016). In our data, conscientiousness and other traits are also associated with diminished noise, suggesting that both cognition and personality influence the quality and stability with which preferences are expressed. We further show that personality traits directly affect guessing behavior and also indirectly through their impact on cognitive noise, which in turn affects guessing. These are new findings in the literature. Figure 1 summarizes our conceptual framework.

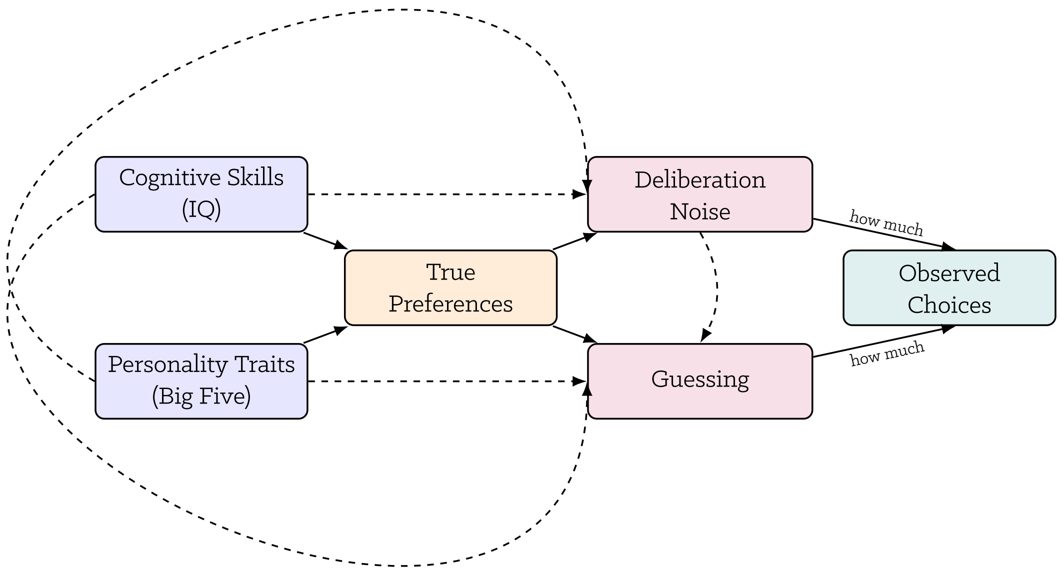


Figure 1: Schemata of Agent Responses to Choice Menus

In our framework, cognitive skills and personality traits jointly shape underlying preferences, while two channels–deliberation noisea and guessing– mediate how these preferences translate into observed choices. 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 improves our understanding of the relationship between choices under risk and psychological traits.

We extend the literature by studying the correlation of decision errors across lists of choices designed to elicit preferences. Independence of preference shocks is typically assumed. We find that the high correlation in shocks is consistent with persistent individual traits or systematic defects in decisions across choices. In contrast, the absence of correlation is consistent with transitory shocks to preferences, as assumed in standard models of noisy cognition. We conduct a parallel analysis for guessing behavior, allowing the variances of shocks affecting guessing to depend on personality, cognitive traits and for shocks to be correlated across items in choice lists.

We quantify the contributions of personality and cognition to: (1) true preference parameters, (2) guessing behavior, (3) transitory shocks to preferences and guessing, and (4) persistent components of decision quality.

We test and reject the standard CRRA specification used in all previous work and find that the Expo-Power utility model provides a superior fit across all of our modifications of the standard model.[3] The choice of functional form critically affects our estimates of risk aversion.

We find strong gender differences in preferences and in the role played by personality traits in each component of our model. Environments influence choices but 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 overturn the consensus that girls are more risk-averse than boys, a pattern that can indeed be replicated in our data when using standard CRRA models without a correlated error structure, but disappears once we account for correlated errors, and further reverses when we relax the CRRA functional-form assumptions. This reversal reveals fundamental limitations in how preferences are measured using conventional methods.

[3]Our evidence on expo-power is consistent with the work of Holt and Laury (2002).

Properly accounting for correlated shocks eliminates it. Flexible functional forms combined with correlated shocks reveal boys are 29% more risk-averse, contradicting decades of research (Croson and Gneezy, 2009; Sutter et al., 2019). Existing work on personality and preferences (Borghans et al., 2008; Dohmen et al., 2011; Jagelka, 2024) emphasizes direct trait-preference links.[4] We show that traits shape decision quality through cognitive noise and guessing behavior.

Second, we demonstrate that standard models systematically misattribute behavioral variation when decision quality varies across individuals. When the various components 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, and allowing deliberation and guessing errors to be correlated. These relaxations of assumptions regarding the error structures have substantial impacts on the estimation of risk preferences, particularly for boys. An additional discovery is the importance of using teacher assessments of personality instead of widely used self-assessments by children. Model fit is better and economic estimates are more interpretable. This is consistent with research by Feng et al. (2022).

The model developed in this paper is an empirically-operational and novel framework for analyzing relationships between preferences and traits. It can be applied to analyze a variety of preference and trait data sets.

The rest of the paper is organized as follows. Section 2 presents a framework for

[4] Andersen et al. (2008) shows that the IQ affects cognitive noise. We extend his 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 cognitive noise.

analyzing choice under cognitive noise and guessing, extending previous models by incorporating personality traits and allowing shocks to preferences to influence guessing behavior. Section 3 describes our data, emphasizing its unique features. Section 4 compares the fit of standard CRRA models with Expo-Power utility specifications, see Xie (2000), which is more flexible and captures wealth-dependent risk aversion and responses of risk choices to environmental variables. Section 5 presents our main empirical results. We decompose the variance in deliberation noise and guessing into persistent individual traits and transitory choice-specific shocks. We estimate how cognitive skills 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 misunderstand how much they truly value one alternative over another or be influenced by transient cognitive or affective factors.[5]

[5] 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; Benjamin et al., 2018; Woodford, 2020; Caplin, 2025). Structural and experimental evidence shows that measured decision imprecision (noise) affects revealed preferences, and that personality can be related to preference parameters in models that handle decision errors (Jagelka, 2024).

2.1 Decision under Imperfect Knowledge

Consider a binary choice between alternatives A and B. Let U^A and U^B denote *true* utilities. Utilities may not be perfectly perceived, but subject to random 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 distortions. For convenience, define $\varepsilon \equiv \tau^B - \tau^A$, and assume $\tau^A \perp\!\!\!\perp \tau^B$.[6] Thus ε summarizes uncertainty about $U^A - U^B$. A positive ε favors B, a negative one favors A. If $\mathrm{Var}(\tau^A) = \sigma^2_{\tau^A}$ and $\mathrm{Var}(\tau^B) = \sigma^2_{\tau^B}$, $\mathrm{Var}(\varepsilon) = \sigma^2_\varepsilon = \sigma^2_{\tau^A} + \sigma^2_{\tau^B}$. A small σ^2_ε implies more stable choices, large σ^2_ε implies near-random choices. Normalizing by σ_ε gives

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

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

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

The gap between (2) and (3) measures decision imperfection.

[6]The symbol "$\perp\!\!\!\perp$" denotes statistical independence. This assumption is not strictly required for the main ideas of this paper.

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. More effort increases precision. $\sigma^2_\tau(\mathcal{I}) \downarrow$ as $\mathcal{I} \uparrow$, for $j = A, B$.

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 costs. Since $D_A + D_B = 1$, realized utility is $U^A D_A + U^B D_B$. Agents are assumed to choose $(\mathcal{I}_A, \mathcal{I}_B)$ to maximize expected utility net of costs:

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

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.

2.3 A Simple Example

Suppose $\varepsilon \sim \mathcal{N}(0, \sigma^2_\varepsilon)$ and define $q \equiv \varepsilon/\sigma_\varepsilon \sim \mathcal{N}(0,1)$. Then the expected payoff (ignoring costs) can be expressed using $\Phi(\cdot)$, the standard normal CDF. When the cost of information acquisition exceeds its marginal benefits, or when attention somehow collapses, $\sigma_\varepsilon(I) \to \infty$ and choices approach a 50–50 randomization—i.e., guessing. Thus,

$$\Pr(D_A = 1) = \Phi\left(\frac{U_A - U_B}{\sigma_\varepsilon}\right), \tag{5}$$

which tends to $\frac{1}{2}$ as $\sigma_\varepsilon \to \infty$.

2.4 Stochastic Specifications

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

$$\varepsilon_{ij}^D = \lambda_j \, \theta_i^D + u_{ij}^D, \tag{6}$$

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

$$\theta_i^D \perp\!\!\!\perp u_{ij}^D, \qquad u_{ij}^D \sim \mathcal{N}(0, \sigma^2_{u_{ij}^D}), \tag{7}$$

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 choice tasks, a central feature of one version of the Thurstone (1927) model. Assuming exchangeability among tasks, we set $\lambda_j \equiv \lambda$, and normalize $\lambda = 1$. We parameterize $\sigma^2_{u_{ij}^D}$ and $\sigma^2_{\theta 0}$ to depend on psychological traits and background variables for each i in $i = 1, \ldots, I$.

2.4.1 Guessing

Before stating a preference, an individual may decide whether or not to exert cognitive effort. If effort is not exerted, the individual randomizes. If it is exerted, they may follow a utility-based rule, as described in (Woodford, 2020; Belzil and Jagelka, 2024; Caplin, 2025). The error structure for the guessing shock ε_{ij}^G for person i on item j parallels that for the decision-noise shock in equation (6):

$$\varepsilon_{ij}^{G} = \omega_j \, \theta_i^{G} + u_{ij}^{G}, \qquad \theta_i^{G} \perp\!\!\!\perp u_{ij}^{G}, \quad u_{ij}^{G} \sim \mathcal{N}(0, \sigma^2_{u_{ij}^{G}}), \tag{8}$$

where θ_i^G represents a persistent (trait-like) disengagement or 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:

$$G_{ij} = \mathbb{1}\Big(\eta(\sigma^2_{u_i^D}, \boldsymbol{X}_i) + \gamma_j \, \theta_i^G + u_{ij}^G > 0\Big), \tag{9}$$

where $\eta(\sigma^2_{u_i^D}, \boldsymbol{X}_i)$ is a function of observed variables capturing how guessing depends on the vector of psychological traits, as well as the structural variance of deliberation noise, $\boldsymbol{X}_i$ is background variables, IQ and personality traits. We consider three parameterizations of the guessing propensity: (i) one in which guessing depends directly on cognitive ability and personality traits, (ii) a second in which deliberation noise σ_{ij}^D affects guessing decisions, with cognitive ability and personality traits influencing guessing only indirectly through deliberation noise, and (iii) a third specification that incorporates both channels described in the previous two models. We use the first specification as the default in our full model 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 that both ε_{ij}^D and ε_{ij}^G exhibit within-choice correlation structures as previously described. Standard random utility models assume choice errors are independent across tasks. However, children appear to exhibit persistent individual differences in decision quality. Some consistently make precise choices, while others persistently exhibit noisy behavior.

2.5 Probabilities

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 is keeping conditioning on X_i implicit to simplify notation, is

$$\Pr(Y_{ij} = B) = \underbrace{\Pr(Y_{ij} = B \mid G_{ij} = 0)}_{\text{deliberate}} \Pr(G_{ij} = 0) + \underbrace{\tfrac{1}{2}}_{\text{guess}} \Pr(G_{ij} = 1). \tag{10}$$

Under normality,

$$\Pr(Y_{ij} = B) = \Phi\left(\frac{U_{ij}^B - U_{ij}^A}{\sigma_{\varepsilon_{ij}}}\right) \left[1 - \Phi(g_{ij})\right] + \frac{1}{2}\,\Phi(g_{ij}), \tag{11}$$

Equation (11) cleanly separates preference-driven choice from random guessing. We assume that

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

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

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

We make a parallel assumption for guessing shocks,

$$\text{Var}[(\varepsilon_{i1}^G \ldots \varepsilon_{iJ}^G)'] = \sigma_{\theta^G}^2 \, (\boldsymbol{i}_J \boldsymbol{i}'_J) + \sigma_{u^G}^2 \, \boldsymbol{I_{JJ}}, \tag{14}$$

with

$$ICC^G = \frac{\sigma_{\theta^G}^2}{\sigma_{\theta^G}^2 + \sigma_{u^G}^2}. \tag{15}$$

2.6 The Log-Likelihood of the Model

The model specifies the probability of choice for each item, generated by latent random variables that govern deliberation and guessing.

Define

$$\varepsilon^G = \begin{pmatrix} \varepsilon_{i1}^G \\ \varepsilon_{i2}^G \\ \vdots \\ \varepsilon_{iJ}^G \end{pmatrix} = \theta_i^G \boldsymbol{i} \quad + \underbrace{\begin{pmatrix} u_{i1}^G \\ u_{i2}^G \\ \vdots \\ u_{iJ}^G \end{pmatrix}}_{\boldsymbol{U}^G}, \qquad \varepsilon^D = \begin{pmatrix} \varepsilon_{i1}^D \\ \varepsilon_{i2}^D \\ \vdots \\ \varepsilon_{iJ}^D \end{pmatrix} = \theta_i^D \boldsymbol{i} \quad + \underbrace{\begin{pmatrix} u_{i1}^D \\ u_{i2}^D \\ \vdots \\ u_{iJ}^D \end{pmatrix}}_{\boldsymbol{U}^D}. \tag{16}$$

where $\boldsymbol{i}$ is a $J \times 1$ vector of ones. Array all model parameters in a matrix Ω. The log-likelihood function is given by the integral of the joint probability over the latent random effects (θ^D, θ^G):

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\Omega}) = \sum_{i=1}^{I} \ln \int\int \prod_{j=1}^{J} \Bigg\{ & \Big[\Pr(Y_{ij} = B \mid \theta^D, \boldsymbol{X}_i, G_{ij} = 0) \Pr(G_{ij} = 0 \mid \boldsymbol{X}_i, \theta^G, \eta_{ij}) \\
& + \tfrac{1}{2} \Pr(G_{ij} = 1 \mid \boldsymbol{X}_i, \theta^G, \eta_{ij}) \Big]^{y_{ij}} \\
& \times \Big[\Pr(Y_{ij} = A \mid \theta^D, \boldsymbol{X}_i, G_{ij} = 0) \Pr(G_{ij} = 0 \mid \boldsymbol{X}_i, \theta^G, \eta_{ij}) \\
& + \tfrac{1}{2} \Pr(G_{ij} = 1 \mid \boldsymbol{X}_i, \theta^G, \eta_{ij}) \Big]^{1-y_{ij}} \Bigg\} \times dF(\theta^D, \theta^G), \qquad (17)
\end{aligned}
$$

where $\boldsymbol{X}_i$ is a vector of background characteristics, IQ and personality traits, and G_{ij} is an indicator of guessing behavior as defined by Equation 9.

$F(\theta^D, \theta^G)$ denotes the joint distribution of the persistent permanent random variables, assumed to be normal and associated with ICC coefficients. These may be correlated. This specification explicitly accounts for the possibility that individuals with higher latent decision precision may also differ systematically in their propensity to guess. The variables θ^D and θ^G are integrated out in forming the probability, and persistent components may be correlated. We allow θ^D and θ^G to be correlated, and parameterize the idiosyncratic and persistent variances $\sigma^2_{u^D}$ and $\sigma^2_{u^G}$ as functions of individual background, cognitive ability, and personality traits, while the elements of u^D and u^G are assumbed to be mutually uncorrelated. These heteroskedastic structures capture how unobserved behavioral noise varies across individuals and tasks. We assume 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 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 function), (iii) the standard deviation of the transitory shock ($\sigma^2_{u^D}$), (iv) the ratio between the standard deviations of the persistent and transitory components of the deliberation noise ($\sigma^2_{\theta^D}/\sigma^2_{u^D}$), (v) the ratio between the standard deviations of the persistent and transitory components of the guessing noise ($\sigma^2_{\theta^G}/\sigma^2_{u^G}$), with $\sigma^2_{u^G}$ normalized to 1, (vi) the correlation between the transitory components of deliberation and guessing noises, $\rho(\theta^D, \theta^G)$, and (vii) the guessing index η as in equation 9.[7]

2.7 Identification

Apesteguia and Ballester (2018) show that models without heteroskedasticity in error variances, guessing behavior, and correlation across choices may generate non-monotonic choice probabilities with respect to risk aversion. Multiple values of a CRRA parameter may be consistent with the same probability of choice for items in choice lists. This possibility leads them and Jagelka (2024) to adopt a random parameter choice model following Thurstone (1927) and Domencich and McFadden (1975) to avoid this possibility. This is unfortunate because Apesteguia and Ballester (2018)'s analysis is irrelevant for identification. Moreover, Jagelka's specification assumes parameters of preference functions associated with observables fluctuate randomly across choices, item by item. We avoid making this strong assumption in our specification and achieve at least local identification.

[7]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.

Non-monotonicity and lack of identification are distinct issues. Identification is a local property of the likelihood function requiring sufficient curvature for unique parameter recovery, not global monotonicity of item mappings in terms of a risk aversion parameter. Our model extends Jagelka (2024) in multiple ways. The vector of J utility contrasts for person i, corresponding to the J items in a choice list, can be written as

$$\boldsymbol{\Delta}_i(\boldsymbol{X}_i, \boldsymbol{\varepsilon}_i^D) \quad \text{, where } \boldsymbol{\varepsilon}_i^D = (\varepsilon_{i1}^D \dots \varepsilon_{iJ}^D)' \tag{18}$$

Jagelka (2024) assumes that the components of $\boldsymbol{\varepsilon}_i^D$ for row j and j' where $j \neq j'$, are independent conditional on $\boldsymbol{X}_i$. We allow for dependence. Both papers allow for heteroskedasticity in the vector $\boldsymbol{\varepsilon}_i^D$ across j but of different functional forms.

We further discuss identification in Section 4.1 after we describe the data and its rich variation in choice probabilities, stakes, and background variables.

3 The Mianzhu Data

Our empirical analysis is based on 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 stands out for three features critical to our analysis. First, we simultaneously measures economic preferences (risk, ambiguity, time, altruism, reciprocity), cognitive ability (IQ), and personality traits (Big Five) in a large and socioeconomically diverse population. Second, we conduct repeated preference elicitations over time, allowing us to separate persistent traits from transitory shocks. Third, we integrate multi-source psychological assessments from students, teachers, and parents, mitigating self-report bias that often contaminates personality–behavior relationships.[8]

In this paper, we focus only on 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 who have observed students over multiple grades. Risk preferences were elicited using six incentivized multiple price list (MPL) tasks administered via tablet computers. 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 sample mean and standard deviation within each grade level. Personality traits were assessed using the Big Five Inventory-2 (BFI-2), consisting of 60 items (12 per dimension) capturing conscientiousness, agreeableness, extraversion, openness to experience, and emotional stability (also known as neuroticism, when reversed) via a validated Chinese translation. Complete details on IQ measurement, personality traits, and their psychometric properties are provided in Appendix B.

[8]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 (2019) and Feng et al. (2022)

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 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) 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

In light of the analysis of Feng et al. (2022), 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. Appendix A validates this choice: while structural parameters remain stable within 5–10% across measurement sources, teacher reports yield 50–100% narrower confidence intervals for trait-preference relationships with identical risk aversion estimates (within 5%). Figure B1 illustrates systematic divergence between teacher and student self-reports across grades, with self-reports consistently higher, particularly for Conscientiousness and Agreeableness. Table B2 shows modest but positive correlations, with stronger agreement for observable behaviors than internal emotional states.

Table 1: Descriptive Statistics: Grade 5 in 2019

	Boys			**Girls**		
	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.
IQ and personality traits						
IQ Index (Standardized)*	1115	−0.014	1.037	1089	0.014	0.961
Conscientiousness	1115	0.246	0.692	1089	0.631	0.651
Agreeableness	1115	0.581	0.866	1089	0.837	0.802
Extraversion	1115	0.066	0.767	1089	0.064	0.818
Openness to Experiences	1115	0.437	0.682	1089	0.454	0.646
Emotional Stability**	1115	0.282	0.750	1089	0.461	0.697
Demographics						
Age	1111	10.707	0.350	1086	10.655	0.316
Living in Rural Area	1115	0.227	0.419	1089	0.233	0.423
Parental Education						
Father Went to High School	1115	0.305	0.461	1089	0.348	0.477
Mother Went to High School	1115	0.249	0.433	1089	0.288	0.453
Family Structure and Migration Status						
Having Any Sibling	1115	0.366	0.482	1089	0.413	0.493
Left-Behind by Father Only	1109	0.250	0.333	1085	0.253	0.330
Left-Behind by Mother Only	1109	0.069	0.185	1085	0.056	0.164
Left-Behind by Both	1109	0.170	0.302	1085	0.154	0.286
LBC Volatility Index	1109	1.454	1.769	1085	1.423	1.731

Notes: Descriptive statistics for Grade 5 students surveyed in 2019. The standardized IQ index is estimated using an IRT-3PL model and standardized to mean 0 and standard deviation 1 within grade. Big Five traits are predicted using the factor model in Heckman et al. (2013), with estimation-error corrections via Bartlett GLS estimator. For each trait, the sample mean is normalized by fixing the "Neutral" response to the first item to zero, standard deviation is normalized by fixing the factor loading of the first item to one. 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. *These scores are standardized relative to the pooled sample of boys and girls, not against a benchmark national sample of fifth graders, which to our knowledge, does not exist. **This trait, when reverse-coded, is also known as Neuroticism in the Big Five Inventory.

3.2 Gender Differences in Traits and Cognitive Development

Substantial gender differences emerge in personality traits. Girls demonstrate markedly higher Conscientiousness (0.631 vs. 0.246) and Agreeableness (0.837 vs. 0.581)—patterns consistent with developmental psychology literature and that persist across Grades 4–6 (Figure B5). Extraversion levels are comparable between genders, while girls hold modest advantages in Openness and Emotional Stability.

Cognitive development follows a convergent trajectory (Figure B5). Although boys exhibit higher IQ scores in Grade 4, girls' faster growth rates produce convergence by Grade 5 and a slight female advantage by Grade 6. Family characteristics reveal that 37% of boys and 41% of girls have siblings. Parental labor migration affects a substantial share of children: on average across semesters, roughly 25% of children experience father-only migration, 6% mother-only migration, and 15% both-parent migration. Children experience approximately 1.4 parental migration status changes on average since starting Grade 1.

3.3 Socioeconomic Gradients and Environmental Influences

The sample displays considerable socioeconomic diversity. Approximately 23% of students live in rural 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%, 6%, and 15% of boys are left behind by their father only, mother only, and both parents, respectively. Girls exhibit similar patterns across these categories. The last row of the table indicates that, on average, boys and girls each experience about 1.4 changes in their parents' migration status since Grade 1.

Figures B2 and B3 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.

Figure B4 reveals substantial rural-urban disparities in cognitive development, with urban students consistently scoring higher and the gap widening from approximately 0.3 standard deviations in Grade 4 to nearly 0.5 by Grade 6. Personality trait differences between rural and urban students are modest, suggesting that environmental factors operate primarily through cognitive rather than socio-emotional development channels.

3.4 Correlations Among Cognitive and Personality Measures

Tables C1, C2, and C3 show moderate positive correlations among cognitive and non-cognitive measures. IQ correlates positively with all Big Five traits, strongest with Openness ($r = 0.24$). Among personality traits, Conscientiousness shows strong correlations with Agreeableness ($r = 0.64$) and Emotional Stability ($r = 0.55$). The correlation structure is largely consistent across gender, though girls exhibit somewhat stronger correlations among the Big Five. These moderate correlations suggest that cognitive ability and personality traits capture distinct sources of individual heterogeneity relevant for decision-making.

3.5 Determinants of Cognitive and Personality Traits

Tables C4 and C5 examine how family background and migration status predict the individual characteristics described in Table 1. Urban residence has mixed effects: urban boys score higher on IQ and Agreeableness but lower on Emotional Stability, while urban girls show substantially lower scores across most personality dimensions—patterns that differ from unconditional comparisons in Figure B4.

Parental education shows positive effects. Father's high school education significantly increases Conscientiousness, Agreeableness, and Openness for boys, with similar patterns for girls. Mother's education is associated with Conscientiousness and Agreeableness for boys, and Conscientiousness, Agreeableness, and Openness for girls, consistent with developmental gradients in Figures B2 and B3.

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 and Emotional Stability. For girls, being left behind by the mother is associated with lower Openness and marginally lower Emotional Stability, while left-behind volatility increases Extraversion.

The relationships between family background and individual traits provide important context for interpreting structural estimates. As we demonstrate in Section 5, differences in risk preferences and decision-making quality across demographic groups operate mainly through their effects on cognitive ability and personality traits, which in turn shape both true preferences and the precision of decision-making[9], as well as

[9]Appendix L demonstrates that cognitive/personality traits and socioeconomic background capture largely independent sources of heterogeneity: structural parameters remain stable (within 1–10%) whether we include SES only, IQ/traits only, or both simultaneously, validating our inclusive

the probability of guessing.

3.6 Some Evidence on Decision Consistency

Before presenting our structural estimates, we examine simple statistical evidence on decision consistency. Table C7 presents linear probability models where the dependent variable measures switching behavior in multiple price lists, where agents oscillate between safe and risky choices across items in the lists as risk increases monotonically. This switching behavior is a source of identifiability for our model.

Several patterns foreshadow our structural findings. IQ significantly reduces inconsistent behavior for both boys and girls, in accordance with the notion that higher cognitive ability improves task comprehension and decision consistency. Among personality traits, Conscientiousness reduces inconsistencies for girls, while Openness strongly reduces it for boys. Interestingly, Extraversion increases inconsistencies for boys but reduces it 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 disrupt decision quality. These reduced-form patterns motivate our structural approach, which disentangles how IQ, personality, and background operate through true risk preferences, cognitive precision in deliberation, guessing behavior, and serial correlation in preference shocks.

modeling approach.

4 Utility Functions: CRRA versus Expo–Power

A widely used benchmark for modeling preferences under risk is the constant relative risk aversion (CRRA) utility function (the consumption level before participation): 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 (2024)

$$U_{\text{CRRA}}(x+w) = \begin{cases} \dfrac{(x+w)^{1-r}-1}{1-r}, & r \neq 1, \\ \\ \ln(x+w), & r = 1, \end{cases} \tag{19}$$

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

$$\text{RRA}(x+w) = r, \tag{20}$$

independently of wealth. We can parametrize r to depend on psychological traits.

A more flexible specification that nests CRRA was introduced by Saha (1993) and Xie (2000). The *expo–power* utility function is:

$$U_{\text{ExpoPower}}(x+w) = \frac{1}{\beta}\left[1 - \exp\left(-\beta\,\frac{(x+w)^{1-r}-1}{1-r}\right)\right], \qquad \beta \geq 0,\ r \geq 0. \tag{21}$$

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

$$\text{RRA}(x+w) = r + \beta\,(x+w)^{1-r}. \tag{22}$$

The parameter r governs baseline curvature, while β scales the sensitivity of risk aversion to the level of $(x + w)$.[10]

When $x = 0$ (i.e., zero payoff from the task), relative risk aversion equals $r + \beta w^{1-r}$, meaning that both parameters jointly determine effective risk aversion at baseline wealth. This specification allows risk aversion to vary with total resources: when $r < 1$, individuals exhibit decreasing relative risk aversion (DRRA) as $(x + w)$ increases if $\beta > 0$, and increasing relative risk aversion (IRRA) if $\beta < 0$. The expo–power specification also bounds utility from above when $\beta > 0$, introducing a natural notion of satiation. Additionally, it exhibits desirable concavity properties under the condition $\beta(x + w)^{1-r} > r - 1$ (see Appendix D for detailed mathematical analysis).

This flexibility enables a better empirical description of behavior in environments where observed heterogeneity in choices may reflect differences in wealth, stakes, or cognitive engagement rather than fixed curvature alone (Xie, 2000).[11] For 10- and 11-year-olds, wealth-dependent risk aversion captures how stake magnitude affects decision-making. Our tasks span 2 to 200 tokens—a 100-fold range that determines prize tier achievement). At low stakes (2 tokens), 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 $\mathrm{RRA}(x+w)$ to vary with this wealth level, capturing psychological shifts across the stake range

[10]For both the CRRA and expo-power models, we assume that baseline wealth is fixed at $w = 1$. In Appendix M, we relax this assumption by fixing w at alternative values. In Appendix N, we relax it further by estimating w as a free parameter and allowing it to vary with children's socioeconomic status.

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

rather than imposing CRRA's rigid constant-aversion assumption. This flexibility is essential for children whose preference structures remain in development and who may process small versus large gambles qualitatively differently.

4.1 Identification

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 pattern, (iii) cross-individual variation in background and psychological traits, and (iv) functional form restrictions. Preference parameters are identified from switching patterns across tasks. Curvature r is pinned down by probability variation for the same expected outcomes. In particular, the identification of the β parameter in the Expo-Power model, which represents the wealth sensitivity of relative risk aversion, is enabled by the broad range of stake sizes (from 2 to 200 tokens) and by variations in w that we study. Transitory deliberation noise σ_{u^D} is identified from within-person variability that cannot be rationalized by stable preferences. ICC^D and ICC^G are identified from correlations in shocks to estimated true preferences and guessing equations across 60 items. The correlation $\rho = \text{Corr}(\theta^D, \theta^G)$ is identified from correlations of shocks across the deliberation and guessing equations. Guessing probabilities are identified from patterns of choices inconsistent with the additive form of the Thurstone model.

Appendix E sketches a proof of local concavity of the likelihood for the case of uncorrelated shocks. A full proof of local concavity is under development. As an imperfect substitute for a rigorous proof, we report minimum eigenvalues of the Hessian and condition numbers for each model that we estimate.

Table E1 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 10^3 to 10^6 (with the corresponding smallest eigenvalues all above 10^{-3}), indicating a moderately well-conditioned curvature of the likelihood surface and suggesting that comparisons across specifications are numerically meaningful. The minimum eigenvalues are bounded away from zero—a computational form of identification.

5 Empirical Results

This section reports empirical results from the various specifications of the model previously described. First, we present the estimate for CRRA without adding any psychological parameters, background variables, or accounting for correlated random shocks, or accounting for guessing behavior. This gives a benchmark model that even in this stark form favors the Expo-Power specification, especially for girls. We then examine the performance of the models as various features are added.

In all of these specifications, we condition on psychological variables and examine their predictive power for each component of each modification we present. We then summarize the results after presenting them. The main message is that accounting for correlated shocks to true preferences and to guessing greatly affects the performance of the model. Psychological measures affect deliberation noise, guessing, and correlation across shocks but do not play a strong role in shaping true preferences. Background affects measured choices through its effect on psychological traits. Conditioning on those traits, experience of the child plays no important role. We also examine variations on equation (9) such as including deliberation noise into the guessing equation. We find that the effect of deliberation noise is strong for girls but weaker for boys.

5.1 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 progressive model building.

Table 2 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. Cognitive 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 decision quality and functional form restrictions.

Table 2: Baseline CRRA Model

	Boys	Girls
r (risk aversion)	0.845***	0.946***
	(0.014)	(0.013)
σ_{u^D} (cognitive 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.

5.2 Parameterization: How Traits Enter the Model

Before presenting the full model results, we describe how cognitive ability and personality traits enter the structural parameters. This discussion is essential for understanding the coefficient estimates reported below.

5.2.1 Parameters and Their Dependence on Observables

The model involves seven key parameters, each of which we allow to depend on individual characteristics $\boldsymbol{X}_i$ (IQ, personality traits, and background variables).

Risk aversion curvature $r(\boldsymbol{X}_i)$ captures baseline curvature of the utility function, where higher r implies greater risk aversion. The wealth sensitivity $\beta(\boldsymbol{X}_i)$ is the Expo-Power parameter governing how risk aversion varies with wealth. When $\beta > 0$ and $r < 1$, individuals exhibit decreasing 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 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 disengage. 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 riskier choices tend to guess more or less often. Finally, guessing probability $\Pr(G_{ij} = 1 \mid \boldsymbol{X}_i)$ is the probability of random guessing rather than deliberation on a given task. In some specifications, we add deliberation variance, $\sigma^2_{u_D}$, as an additional parameter.

5.2.2 Functional Form of Dependence

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

$$\pi(\boldsymbol{X}_i) = \sum_{k=0}^{K} \pi_k \cdot X_{ki}, \tag{23}$$

where π_k represents the parameters of interest.

For parameters that must be 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:

$$\sigma^2_{u^D}(\boldsymbol{X}_i) = \exp\left(\sum_{k=1}^{K} \gamma_k \cdot X_{ki}\right), \tag{24}$$

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 $(-1, 1)$.

5.2.3 Structural Interpretation

A key distinction from reduced-form approaches is that our coefficients represent effects on structural parameters within a fully specified behavioral model, not associations with choices. When we estimate that Conscientiousness increases r by 0.104 for boys, this means conscientious boys have higher baseline risk aversion—not merely that they make choices that appear safer. The latter could arise through multiple channels (higher precision, less guessing, different preferences). Our structural model disentangles these mechanisms.

Similarly, changes in σ_{u^D} represent changes in decision precision, not choice consistency per se. A trait that reduces σ_{u^D} makes choices more precisely reflect true preferences, reducing measurement error in revealed risk aversion. Effects on the correlation ρ reveal whether traits systematically link deliberation quality and guessing behavior—for example, whether individuals with traits favoring riskier choices tend to guess more (negative ρ) or deliberate more carefully (positive ρ).

Identification of these covariate effects exploits cross-individual variation, as detailed in Section 4.1. Traits affecting preferences shift average switching points across individuals, traits affecting noise increase within-person variability, traits affecting ICC affect cross-task correlations in choice patterns, and traits affecting the correlation ρ affect the relationship between deliberation precision and guessing propensity across individuals.

5.3 Building up to the Full Model

We incrementally add model components to our empirical model to understand their individual and joint contributions.

5.3.1 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 estimated risk aversion 34% lower than under CRRA. For boys, the baseline comparison shows no significant difference (LR = 2.0, $p = 0.157$), though this changes when incorporating heterogeneity.

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

	Boys		Girls	
	CRRA	Expo-Power	CRRA	Expo-Power
	(1)	(2)	(3)	(4)
Risk Mean	0.845***	0.801***	0.946***	0.626***
$R(x,w)$ at $x=0$, $w=1$	(0.014)	(0.038)	(0.013)	(0.029)
r in Expo-Power		0.767***		0.546***
		(0.055)		(0.025)
β in Expo-Power		0.034*		0.080***
		(0.018)		(0.004)
Cognitive 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: 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. $R(x,w)$ denotes relative risk aversion evaluated at $x = 0$ and $w = 1$. For CRRA, $R(x,w) = r$. For Expo-Power, $R(x,w) = r + \beta(x+w)^{1-r}$.

5.3.2 Components of Decision Quality: ICC Structure and Guessing Behavior

A fundamental question is whether decision errors reflect purely transitory 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 decision precision, and explicit guessing behavior that models strategic disengagement. 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. 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 (Chernoff, 1954; Self and Liang, 1987). We use this mixed χ^2 distribution to calculate the p-values for these two likelihood ratio tests.

Table 4 presents specifications incorporating each component separately. Panel A examines 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. Critically, Expo-Power maintains its advantage over CRRA even after both include ICC (boys: LR = 16.2, $p < 0.001$, girls: LR = 182.3, $p < 0.001$), demonstrating that functional form flexibility and proper error structure are complements rather than substitutes.

The cognitive 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 random noise. For Expo-Power with ICC, noise estimates are 1.046 for boys and 1.720 for girls. This effect 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 F demonstrates that, in the full-model setting, omitting ICC^D inflates transitory noise estimates by 90–170% (Figure F4), systematically underestimates risk aversion by up to 19% (Figures F1–F2), and misallocates strategic disengagement to random noise (Figure F6)—confirming that roughly 60% of apparent noise reflects stable individual differences. Appendix G further shows that ICC omission attenuates trait effects on preferences. Without ICC^D, the Conscientiousness coefficient on risk aversion drops from 0.120 ($p < 0.01$) to approximately 0.05 (not significant), demonstrating that proper error structure is essential for recovering trait-preference relationships (Figure G1).

Panel B examines guessing behavior without ICC^D. Beyond cognitive noise in deliberation, children may sometimes disengage entirely and respond randomly. Adding guessing improves model fit (boys: LR = 50.6–57.4, girls: LR = 3.6–182.0), 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%. Notably, guessing behavior matters much more for explaining girls' choices under Expo-Power (LR = 182.0) than under CRRA (LR = 3.6), suggesting that functional form and guessing interact in capturing girls' decision processes.

Expo-Power maintains superiority over CRRA after incorporating guessing (boys: LR = 8.9, $p = 0.003$, girls: LR = 178.5, $p < 0.001$). Appendix H demonstrates that in the full model, guessing captures qualitatively distinct decision processes: models omitting guessing inflate transitory noise by 90–120% (Figure H4), collapse ICC estimates by 55–60%, and systematically underestimate risk aversion by 4–9%

(Figures H1–H2), demonstrating that strategic disengagement operates through both persistent and transitory channels.

Table 4: 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						
Risk Mean	0.845***	0.939***	0.842***	0.946***	0.949***	0.639***
$R(x,w)$ at $x=0$, $w=1$	(0.014)	(0.010)	(0.030)	(0.013)	(0.009)	(0.028)
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***	3929.2***		3535.8***	3718.0***
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
Risk Mean	0.845***	0.843***	0.758***	0.946***	0.948***	0.605***
$R(x,w)$ at $x=0$, $w=1$	(0.014)	(0.015)	(0.034)	(0.013)	(0.014)	(0.028)
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***	57.4***		3.6**	182.0***
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. $R(x,w) = r + \beta(x+w)^{1-r}$ for Expo-Power and $R(x,w) = r$ for CRRA, evaluated at $x = 0$ and $w = 1$. $\text{ICC}^D = \sigma^2_{\theta^D}/(\sigma^2_{u^D} + \sigma^2_{\theta^D})$ measures the proportion of variance in deliberation noise attributable to its persistent component. Due to the boundary constraints in the parameter space for ICC and guessing, the asymptotic p-values of the LR tests for +ICC and +Guess 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). p-values for LR: Expo-Power vs. CRRA are based on a regular χ^2 distribution.

5.3.3 The Full Model

We now 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 unobservables in deliberation, and (3) explicit guessing behavior with correlation of deliberation-guessing in unobserved shocks. This is our preferred model. When background variables are added, they barely change the estimates. Any effect of background variables on choice parameters operates through psychological variables, see Section 5.8.1.

Table 5 traces how risk aversion estimates evolve as we move from the baseline specification (CRRA without ICC, guessing, or traits) to the full model. The comparison reveals two distinct channels through which modeling choices affect gender gaps in measured risk aversion. Complete coefficient estimates appear in Tables K1–K2 in Appendix K.

The first channel operates through decision quality. 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 covariate effects while maintaining the CRRA functional form. For boys, this transition increases estimated risk aversion from $R(x, w) = 0.845$ to 0.970—a 15% increase. For girls, the increase is smaller: from 0.946 to 0.970, just 3%. This asymmetric response to decision quality controls narrows the apparent gender gap: under baseline CRRA, girls appear 12% more risk-averse than boys. Under full CRRA, the gap disappears entirely.

The second channel operates through use of a more flexible functional form. Moving from full CRRA to full Expo-Power (columns 2 vs. 3 for boys, columns 5 vs. 6 for girls) allows wealth-dependent risk aversion while holding decision quality constant. For boys, this reduces estimated risk aversion modestly from 0.970 to 0.862—an 11% decrease. For girls, the reduction is dramatic: from 0.970 to 0.669, a 31% decrease. This asymmetric response to functional form flexibility reverses the gender gap: under full Expo-Power, boys are 29% more risk-averse than girls.

Cognitive noise shows dramatic reductions when heterogeneity is modeled. For boys, σ_{u^D} drops from 1.368 (baseline) to 0.461 (full CRRA) to 0.556 (full Expo-Power). For girls, σ_{u^D} falls from 1.056 to 0.509 to 0.888. 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 decision quality measurement.

The ICC^D parameters reveal that approximately 40% of decision variance reflects stable individual traits. For deliberation, ICC^D equals 0.432 (boys) and 0.375 (girls). For guessing, ICC^G equals 0.567 (boys) and 0.536 (girls). The higher ICC in guessing suggests strategic disengagement is more trait-like than decision precision. Boys guess on approximately 13–14% of trials, and girls guess on 16%.

The log-likelihood improvements confirm that both channels matter substantially. Moving from baseline to full CRRA improves fit by 3,238 points for boys and 2,677 points for girls. The subsequent move from full CRRA to full Expo-Power yields additional improvements of 22 points for boys but 103 points for girls—reflecting a stronger preference for Expo-Power among girls (LR test: $\chi^2(7) = 44.7$ for boys vs. $\chi^2(7) = 206.4$ for girls, both $p < 0.001$). This confirms that wealth-dependent risk aversion is important but less critical for boys.

Table 5: 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
Risk Mean	0.845***	0.970***	0.862***	0.946***	0.970***	0.669***
$R(x, w)$ at $x = 0$, $w = 1$	(0.014)	(0.012)	(0.028)	(0.013)	(0.011)	(0.028)
σ_{u^D}	1.368***	0.461***	0.556***	1.056***	0.509***	0.888***
	(0.073)	(0.026)	(0.041)	(0.053)	(0.026)	(0.071)
ICC^D	—	0.432***	0.426***	—	0.375***	0.377***
		(0.025)	(0.035)		(0.019)	(0.019)
ICC^G	—	0.567***	0.580***	—	0.536***	0.534***
		(0.024)	(0.027)		(0.023)	(0.025)
$\rho(\theta^D, \theta^G)$	—	−0.144***	−0.151***	—	0.101*	0.115
		(0.036)	(0.034)		(0.057)	(0.074)
Mean Prob(Guess)	—	0.137***	0.127***	—	0.162***	0.160***
		(0.015)	(0.014)		(0.014)	(0.017)
Loglikelihood	−36015.707	−32777.388	−32755.039	−36012.605	−33335.742	−33232.537
Δ vs. Baseline	—	3238.3	3260.7	—	2676.9	2780.1
LR: Expo-Power vs. CRRA			44.698***			206.41***
p-value			[0.000]			[0.000]
AIC	72035.414	65638.776	65608.078	72029.210	66755.484	66563.074

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. $R(x, w) = r + \beta(x + w)^{1-r}$ for Expo-Power and $R(x, w) = r$ for CRRA, evaluated at $x = 0$, $w = 1$, and sample mean trait values. LR test for Expo-Power vs. CRRA has 7 degrees of freedom (one β parameter plus six trait interactions).

5.3.4 Sensitivity to the Components of the Full Model

Table 6 documents how model fit and parameter estimates evolve as components are added. Panel A shows that each component contributes significantly to model fit, with ICC structure providing the largest improvements (approximately 1,950–1,960 log-likelihood points for boys, 1,770–1,860 for girls). Panel B confirms that Expo-Power dominates CRRA by an AIC criterion in the full specification. Panel C reveals that risk aversion estimates are sensitive to model specification: for girls, CRRA estimates remain stable around 0.95 regardless of specification, while Expo-Power estimates range from 0.605 to 0.669 depending on which components are included.

Table 6: Sensitivity to Alternative Specifications: CRRA vs. Expo-Power Across 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	−35662.352	−35631.043	62.618***	−35741.41	−35643.578	195.664***
			(0.000)			(0.000)
Only ICC	−34059.224	−34051.104	16.24***	−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	−32777.388	−32755.039	44.698***	−33335.742	−33232.537	206.41***
			(0.000)			(0.000)
Panel B: Akaike Information Criterion (AIC)						
Baseline (None)	72035.414	72035.414		72029.21	71876.584	
Only IQ and Traits	71352.704	71304.086		71510.82	71329.156	
Only ICC	68124.448	68110.208		68495.414	68315.162	
Only Guessing	71986.858	71979.964		72027.588	71851.116	
Full Model	65638.776	65608.078		66755.484	66563.074	
Panel C: Implied Risk Aversion $R(x, w)$ at $x = 0$, $w = 1$						
Baseline (None)	0.845	0.801		0.946	0.626	
	(0.014)	(0.038)		(0.013)	(0.029)	
Only IQ and Traits	0.842	0.733		0.950	0.659	
	(0.015)	(0.094)		(0.014)	(0.037)	
Only ICC	0.939	0.842		0.949	0.639	
	(0.010)	(0.030)		(0.009)	(0.028)	
Only Guessing	0.843	0.758		0.948	0.605	
	(0.015)	(0.034)		(0.014)	(0.028)	
Full Model	0.970	0.862		0.970	0.669	
	(0.012)	(0.028)		(0.011)	(0.028)	

Notes: Standard errors in parentheses are clustered at the individual level for Panel C. 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 Akaike Information Criterion, where lower values indicate better fit. "Baseline (None)" is the simplest model without IQ/traits, ICC, or guessing. Panel C shows estimated relative risk aversion at $x = 0$ and $w = 1$, where $R(x, w) = r$ for CRRA and $R(x, w) = r + \beta(x + w)^{1-r}$ for Expo-Power evaluated at sample mean trait values.

5.3.5 Impacts on Estimated Model Parameters

Table 7 traces how Expo-Power parameter estimates evolve as model components are sequentially added. For boys, baseline curvature r increases from 0.767 to 0.770 as ICC and guessing are incorporated, while wealth-sensitivity β roughly triples from 0.034 to 0.091. Cognitive noise drops dramatically from 1.485 to 0.556 (63% reduction), confirming that simpler models conflate decision imprecision with preference heterogeneity. For girls, the pattern differs: r remains relatively stable (0.546 to 0.576), while β increases modestly (0.080 to 0.093). The key difference is that girls' cognitive noise reduction is smaller (56% vs. 63%), and their ICC estimates are lower (0.377 vs. 0.426 for deliberation), suggesting that boys' decision quality is more trait-like.

Table 7: Building Toward the Full Model: Expo-Power Specification

	Model Specification			
Parameter	(1) Baseline	(2) +ICC	(3) +Guess	(4) Full
Panel A: Boys				
r (curvature)	0.767***	0.760***	0.711***	0.770***
	(0.055)	(0.034)	(0.041)	(0.033)
β (wealth dependence)	0.034*	0.082***	0.047***	0.091***
	(0.018)	(0.007)	(0.009)	(0.009)
Risk Mean	0.801***	0.842***	0.758***	0.862***
$R(x, w)$ at $x = 0$, $w = 1$	(0.038)	(0.030)	(0.034)	(0.028)
σ_{u^D} (noise)	1.485***	1.046***	1.227***	0.556***
	(0.128)	(0.072)	(0.098)	(0.041)
ICC^D	—	0.180***	—	0.426***
		(0.010)		(0.035)
ICC^G	—	—	—	0.580***
				(0.027)
$\rho(\theta^D, \theta^G)$	—	—	—	−0.151***
				(0.034)
Pr(Guess)	—	—	0.164***	0.127***
			(0.018)	(0.014)
Panel B: Girls				
r (curvature)	0.546***	0.553***	0.527***	0.576***
	(0.025)	(0.024)	(0.024)	(0.025)
β (wealth dependence)	0.080***	0.086***	0.078***	0.093***
	(0.004)	(0.004)	(0.004)	(0.005)
Risk Mean	0.626***	0.639***	0.605***	0.669***
$R(x, w)$ at $x = 0$, $w = 1$	(0.029)	(0.028)	(0.028)	(0.028)
σ_{u^D} (noise)	2.013***	1.720***	1.678***	0.888***
	(0.163)	(0.134)	(0.141)	(0.071)
ICC^D	—	0.163***	—	0.377***
		(0.009)		(0.019)
ICC^G	—	—	—	0.534***
				(0.025)
$\rho(\theta^D, \theta^G)$	—	—	—	0.115
				(0.074)
Pr(Guess)	—	—	0.143***	0.160***
			(0.024)	(0.017)

Notes: 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. (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 guessing behavior, but not the ICC structure or 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. N = 1,115 for boys and N = 1,089 for girls.

5.3.6 Predicted Values Across Specifications

Tables 8 and 9 provide predicted values at mean IQ and traits across all model specifications.

Table 8 presents results for boys. Several patterns emerge. Risk aversion estimates are consistently higher under CRRA than Expo-Power, with the gap approximately 0.11 in the full specification (0.970 vs. 0.862). Adding IQ and personality traits has modest effects on ICC^D but slightly reduces ICC^G, suggesting that observed traits explain some of the persistent individual differences in deliberation errors. The correlation between the persistent components of deliberation and guessing noise $Corr(\theta^D, \theta^G)$ is consistently negative (-0.15 to -0.19), indicating that boys with unobserved traits favoring riskier choices tend to guess more often. Cognitive noise estimates are highly sensitive to model specification, with the full model yielding estimates 60–70% lower than baseline models.

Table 8: Predicted Values of Model Parameters at Mean IQ and Traits: Boys

	Parameter							
	$R(x,w)$	r	β	σ_{u^D}	ICC^D	ICC^G	$Correl(\theta^D,\theta^G)$	Pr(Guess)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Expo-Power, with IQ and personality traits								
With ICC and Guessing	0.862***	0.770***	0.091***	0.556***	0.426***	0.580***	−0.151***	0.127***
	(0.028)	(0.033)	(0.009)	(0.041)	(0.035)	(0.027)	(0.034)	(0.014)
No ICC, with Guessing	0.705***	0.654***	0.051***	1.496***				0.109***
	(0.052)	(0.058)	(0.008)	(0.209)				(0.027)
With ICC, No Guessing	0.784***	0.704***	0.080***	1.215***	0.186***			
	(0.034)	(0.037)	(0.006)	(0.105)	(0.010)			
No ICC, No Guessing	0.733***	0.689***	0.044**	1.739***				
	(0.094)	(0.112)	(0.019)	(0.378)				
CRRA, with IQ and personality traits								
With ICC and Guessing	0.970***	0.970***		0.461***	0.432***	0.567***	−0.144***	0.137***
	(0.012)	(0.012)		(0.026)	(0.025)	(0.024)	(0.036)	(0.015)
No ICC, with Guessing	0.840***	0.840***		1.217***				0.072***
	(0.015)	(0.015)		(0.082)				(0.023)
With ICC, No Guessing	0.936***	0.936***		0.910***	0.184***			
	(0.011)	(0.011)		(0.042)	(0.010)			
No ICC, No Guessing	0.842***	0.842***		1.394***				
	(0.015)	(0.015)		(0.079)				
Expo-Power, without IQ or personality traits								
With ICC and Guessing	0.876***	0.777***	0.099***	0.519***	0.425***	0.632***	−0.189***	0.126***
	(0.026)	(0.029)	(0.007)	(0.034)	(0.020)	(0.023)	(0.044)	(0.012)
No ICC, with Guessing	0.758***	0.711***	0.047***	1.227***				0.164***
	(0.034)	(0.041)	(0.009)	(0.098)				(0.018)
With ICC, No Guessing	0.842***	0.760***	0.082***	1.046***	0.180***			
	(0.030)	(0.034)	(0.007)	(0.072)	(0.010)			
No ICC, No Guessing	0.801***	0.767***	0.034*	1.485***				
	(0.038)	(0.055)	(0.018)	(0.128)				
CRRA, without IQ or personality traits								
With ICC and Guessing	0.975***	0.975***		0.446***	0.422***	0.626***	−0.177***	0.127***
	(0.012)	(0.012)		(0.022)	(0.020)	(0.023)	(0.043)	(0.012)
No ICC, with Guessing	0.843***	0.843***		1.062***				0.156***
	(0.015)	(0.015)		(0.066)				(0.018)
With ICC, No Guessing	0.939***	0.939***		0.895***	0.178***			
	(0.010)	(0.010)		(0.038)	(0.010)			
No ICC, No Guessing	0.845***	0.845***		1.368***				
	(0.014)	(0.014)		(0.073)				

Notes: Each entry is the prediction at mean IQ and traits. Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. $R(x,w) = r + \beta(x+w)^{1-r}$ for Expo-Power and $R(x,w) = r$ for CRRA, evaluated at $x = 0$ and $w = 1$. N = 1,115.

Table 9 presents the corresponding results for girls. The patterns differ in important ways. The gap between CRRA and Expo-Power risk aversion estimates is much larger for girls (approximately 0.30 in the full specification: 0.970 vs. 0.669), reflecting the binding nature of CRRA's functional form restrictions for girls. Most strikingly, the correlation between persistent components of deliberation and guessing noise is positive for girls (+0.06 to +0.12), opposite to boys. This indicates that girls with unobserved traits favoring riskier choices tend to deliberate more rather than guess—a fundamentally different decision architecture. Cognitive noise under Expo-Power is substantially higher for girls than boys (0.888 vs. 0.556), suggesting that girls' choices exhibit greater variability even after accounting for ICC structure and guessing.

Table 9: Predicted Values at Mean IQ and Traits: Girls

	Parameter							
	$R(x,w)$	r	β	σ_{u^D}	ICC^D	ICC^G	$Correl(\theta^D,\theta^G)$	Pr(Guess)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Expo-Power, with IQ and personality traits								
With ICC and Guessing	0.669***	0.576***	0.093***	0.888***	0.377***	0.534***	0.115	0.160***
	(0.028)	(0.025)	(0.005)	(0.071)	(0.019)	(0.025)	(0.074)	(0.017)
No ICC, with Guessing	0.654***	0.571***	0.083***	1.688***				0.062**
	(0.067)	(0.066)	(0.005)	(0.240)				(0.027)
With ICC, No Guessing	0.645***	0.557***	0.088***	1.720***	0.164***			
	(0.031)	(0.029)	(0.004)	(0.138)	(0.009)			
No ICC, No Guessing	0.659***	0.579***	0.081***	1.901***				
	(0.037)	(0.036)	(0.005)	(0.173)				
CRRA, with IQ and personality traits								
With ICC and Guessing	0.970***	0.970***		0.509***	0.375***	0.536***	0.101*	0.162***
	(0.011)	(0.011)		(0.026)	(0.019)	(0.023)	(0.057)	(0.014)
No ICC, with Guessing	0.953***	0.953***		0.968***				0.031
	(0.014)	(0.014)		(0.059)				(0.020)
With ICC, No Guessing	0.949***	0.949***		0.972***	0.163***			
	(0.009)	(0.009)		(0.038)	(0.009)			
No ICC, No Guessing	0.950***	0.950***		1.052***				
	(0.014)	(0.014)		(0.054)				
Expo-Power, without IQ or personality traits								
With ICC and Guessing	0.676***	0.582***	0.095***	0.850***	0.384***	0.575***	0.056	0.164***
	(0.026)	(0.023)	(0.005)	(0.062)	(0.021)	(0.023)	(0.084)	(0.014)
No ICC, with Guessing	0.605***	0.527***	0.078***	1.678***				0.143***
	(0.028)	(0.024)	(0.004)	(0.141)				(0.024)
With ICC, No Guessing	0.639***	0.553***	0.086***	1.720***	0.163***			
	(0.028)	(0.024)	(0.004)	(0.134)	(0.009)			
No ICC, No Guessing	0.626***	0.546***	0.080***	2.013***				
	(0.029)	(0.025)	(0.004)	(0.163)				
CRRA, without IQ or personality traits								
With ICC and Guessing	0.972***	0.972***		0.500***	0.381***	0.582***	0.078	0.163***
	(0.011)	(0.011)		(0.024)	(0.021)	(0.021)	(0.070)	(0.014)
No ICC, with Guessing	0.948***	0.948***		0.956***				0.067**
	(0.014)	(0.014)		(0.075)				(0.033)
With ICC, No Guessing	0.949***	0.949***		0.965***	0.161***			
	(0.009)	(0.009)		(0.037)	(0.009)			
No ICC, No Guessing	0.946***	0.946***		1.056***				
	(0.013)	(0.013)		(0.053)				

Notes: Each entry is the prediction at mean IQ and traits. Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. $R(x,w) = r + \beta(x+w)^{1-r}$ for Expo-Power and $R(x,w) = r$ for CRRA, evaluated at $x = 0$ and $w = 1$. $N = 1{,}089$.

5.3.7 Comparison Across Studies

Table 10 presents a comparison of risk aversion estimates across studies using incentivized experiments. Panel A focuses on children and adolescents. Our estimates for Chinese children—0.862 for boys and 0.669 for girls—fall within the range observed in other youth samples, though direct comparisons require caution due to differences in utility specifications. Notably, boys in our sample exhibit significantly higher risk aversion than girls. This pattern contrasts with the well-documented finding in adult populations that women tend to be more risk averse than men (Croson and Gneezy, 2009), suggesting that gender differences in risk preferences may emerge, reverse, or develop during the transition from childhood to adulthood. Jagelka (2024), using a CRRA specification with a Random Preference Model, estimates an average risk aversion of 1.01 for Canadian high school students aged 16–18, suggesting that risk aversion may increase through adolescence. Tymula et al. (2012) report that U.S. adolescents (aged 12–17) are actually *more* risk averse than adults (aged 30–50), with estimates of 0.45 and 0.28 respectively. Their finding challenges the conventional view that adolescents are risk-seeking; instead, they argue that adolescent risk-taking behavior stems from tolerance for ambiguity rather than preferences over known risks.

Panel B presents estimates for adult populations. The Danish estimates from Andersen et al. (2008) and Apesteguia and Ballester (2018) cluster in the range of 0.66–0.75, with the latter demonstrating that standard Random Utility Model estimation understates risk aversion by approximately 14% relative to the Random Parameter Model. The highest estimate among adults comes from Tanaka and Munro

(2014), who find a CRRA coefficient of 1.32 among Ugandan farmers—substantially higher than estimates from developed-country samples. This elevated risk aversion likely reflects the greater background risk faced by agricultural households in developing countries, where formal insurance mechanisms are scarce and income volatility is high[12].

Table 10: Mean Risk Aversion Estimates Across Studies

Paper	Sample	Age	N	Utility Specification	Risk Aversion
Panel A: Children and Adolescents					
This paper	Chinese children (Boys)	10.7	1,115	Expo-Power	0.862
This paper	Chinese children (Girls)	10.7	1,089	Expo-Power	0.669
Jagelka (2024)	Canadian high school students	16–18	1,224	CRRA	1.01
Tymula et al. (2012)	U.S. adolescents	12–17	32	Power	0.45
Panel B: Adults					
Tymula et al. (2012)	U.S. adults	30–50	31	Power	0.28
Andersen et al. (2008)	Danish adults	19–75	253	CRRA	0.741
Apesteguia and Ballester (2018)	Danish adults (RPM)	19–75	253	CRRA	0.752
Apesteguia and Ballester (2018)	Danish adults (RUM)	19–75	253	CRRA	0.661
Tanaka and Munro (2014)	Ugandan farmers	45 (mean)	1,279	CRRA	1.32

Notes: Risk aversion estimates from incentivized experiments across studies. For this paper, we report the Risk Mean $R(x, w)$ evaluated at $x = 0$, $w = 1$ from the Expo-Power specification (Full Model, Column 4 in Table 7). Jagelka (2024) estimates CRRA risk aversion using a Random Preference Model with factor structure linking preferences to personality traits. Tymula et al. (2012) report $1 - \alpha$ where α is the power parameter in $U = v^{\alpha}$. Andersen et al. (2008) jointly estimate risk and time preferences using CRRA utility. Apesteguia and Ballester (2018) re-analyze the Andersen et al. (2008) data comparing Random Parameter Model (RPM) and Random Utility Model (RUM) estimators. Tanaka and Munro (2014) estimate CRRA curvature parameter σ from gains-only lotteries. They separately estimate loss aversion ($\lambda = 3.93$) using Prospect Theory. Higher values indicate greater risk aversion in all specifications.

[12] Several methodological considerations warrant mention. First, the utility specifications differ across studies: our Expo-Power specification allows for more flexible curvature than standard CRRA, while Tymula et al. (2012) use a power utility parameterization where $1 - \alpha$ represents risk aversion. Second, sample characteristics vary considerably—from laboratory experiments with university students to large-scale field experiments with representative populations. Third, the stakes and experimental protocols differ, which may affect revealed preferences. Despite these caveats, the broad pattern suggests that children and adolescents exhibit moderate risk aversion comparable to or exceeding that of adults, and that substantial heterogeneity exists across populations and contexts.

5.4 Do IQ and Personality Traits Matter for Economic Preferences?

A central question of this paper is whether individual differences in IQ and personality systematically shape economic preferences and decision quality. We address this by examining covariate effects across all model parameters, using the full Expo-Power specification as our preferred model.

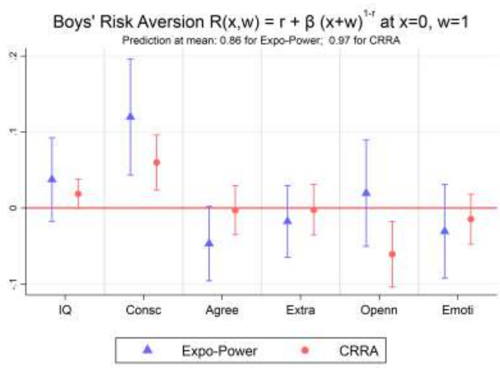
5.4.1 Effects of IQ and Personality Traits on Risk Preferences

Figure 2 presents the effects of IQ and personality traits on overall risk aversion $R(x, w) = r + \beta(x + w)^{1-r}$ evaluated at $x = 0$ and $w = 1$. To compute this, we evaluate their impacts on r and β. The figure compares Expo-Power (blue squares) and CRRA (red circles) specifications.

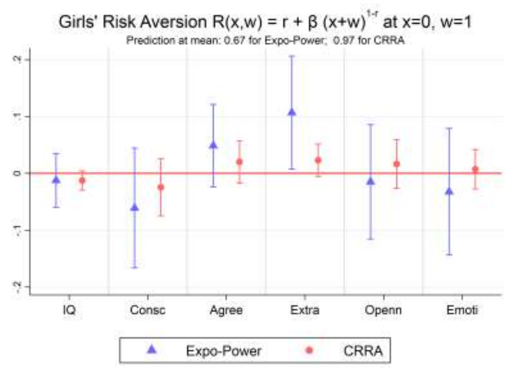
For boys, Conscientiousness significantly increases risk aversion under both specifications (Expo-Power: 0.120, $p < 0.01$, CRRA: 0.152, $p < 0.05$). This finding aligns with theoretical predictions: conscientious individuals, characterized by deliberation and impulse control, may exhibit greater caution in risky choices. Agreeableness shows a marginally negative effect (-0.050, $p < 0.10$), suggesting more agreeable boys are slightly less risk-averse.

For girls, the pattern differs substantially. Extraversion significantly increases risk aversion (Expo-Power: 0.107, $p < 0.05$, CRRA: 0.142, $p < 0.05$). This finding may reflect the social context of our experimental setting, where extraverted girls engage more carefully with the task.

IQ shows no significant direct effects on risk preferences for either gender, suggesting that IQ operates primarily through decision quality rather than preference parameters.



(a) Boys



(b) Girls

Figure 2: Covariate Effects on Risk Aversion: Full Model Comparison

Notes: Coefficient estimates for effects of IQ and Big Five on $R(x, w) = r + \beta(x + w)^{1-r}$ evaluated at $x = 0$ and $w = 1$. Blue squares: Expo-Power, red circles: CRRA. Boys: Conscientiousness increases risk aversion (coef. = 0.120, $p < 0.01$), mean prediction 0.86 (Expo-Power) vs. 0.97 (CRRA). Girls: Extraversion increases risk aversion (coef. = 0.107, $p < 0.05$), mean prediction 0.67 (Expo-Power) vs. 0.97 (CRRA). Error bars show 90% confidence intervals. N = 1,115 boys, N = 1,089 girls.

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

Figure 3 decomposes the effects of IQ and personality traits on overall risk aversion into effects on baseline curvature r (top panels) and wealth-sensitivity 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. = 0.104, $p < 0.01$) with a modest positive effect on β (coef. = 0.016, $p < 0.10$). 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 (-0.043, $p < 0.05$) but no significant effect on r, suggesting that open boys exhibit flatter risk aversion profiles.

For girls, Extraversion increases both r (coef. $= 0.090$, $p < 0.10$) and β (coef. $= 0.017$, $p < 0.10$), indicating that extraverted girls are more risk-averse at all wealth levels with steeper wealth-dependence.

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 $\beta = 0.09$. The gender gap in overall risk aversion ($R(x, w) = 0.86$ vs. 0.67) stems primarily from differences in baseline curvature r, not wealth-sensitivity β.

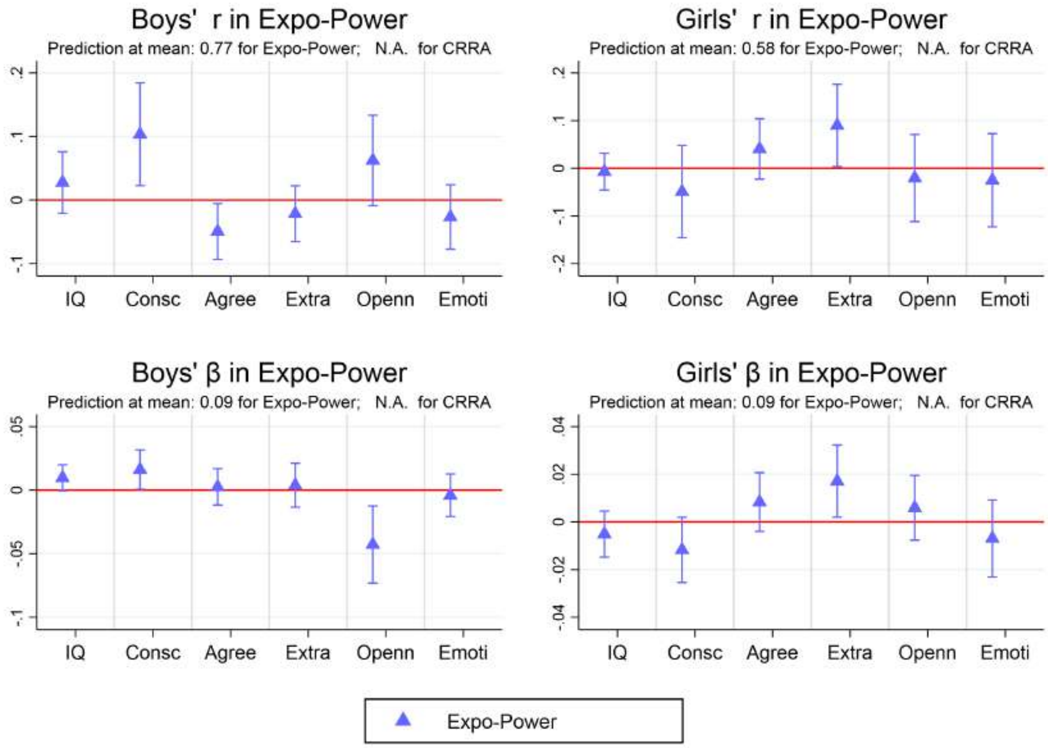


Figure 3: Decomposing Risk Aversion: Effects on r and β Parameters

Notes: Top panels show effects on baseline curvature r, bottom panels show effects on wealth-sensitivity β. Boys: Conscientiousness strongly increases r (coef. $= 0.104$, $p < 0.01$) with modest effect on β (coef. $= 0.016$, $p < 0.10$). Girls: Extraversion increases both r (coef. $= 0.090$, $p < 0.10$) and β (coef. $= 0.017$, $p < 0.10$). Mean predictions—Boys: $r = 0.77$, $\beta = 0.09$, Girls: $r = 0.58$, $\beta = 0.09$. Error bars show 90% confidence intervals.

5.4.3 Effects of IQ and Personality Traits on Decision Quality

Figure 4 presents the effects of IQ and personality traits on transitory noise σ_{u^D} (top panels) and ICC^D (bottom panels). These parameters capture decision quality: lower noise and higher ICC indicate more precise and consistent decision-making.

For boys, Conscientiousness reduces cognitive noise (-0.201, $p < 0.01$) under

Expo-Power, indicating that conscientious boys make more precise decisions. IQ also reduces noise (-0.103, $p < 0.10$), consistent with the interpretation that IQ improves decision precision. Effects on ICC^D are weaker, with Emotional Stability showing a marginally negative effect (-0.079, $p < 0.05$).

For girls, Extraversion reduces cognitive noise (-0.252, 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 $\sigma_{u^D} = 0.56$ (Expo-Power) vs. 0.46 (CRRA), girls show $\sigma_{u^D} = 0.89$ (Expo-Power) vs. 0.51 (CRRA). The larger discrepancy for girls reflects their greater sensitivity to functional form restrictions.

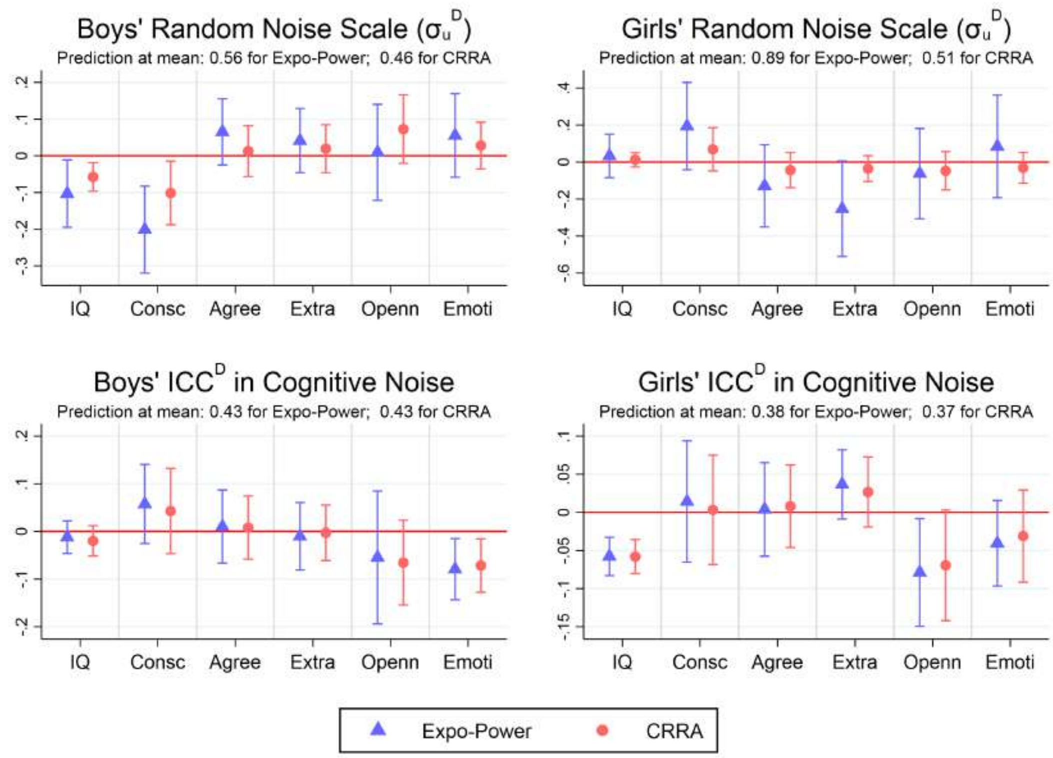


Figure 4: Covariate Effects on Decision Quality: Noise and ICC

Notes: Top panels show effects on transitory noise σ_u^D, bottom panels show effects on ICC^D. Blue squares: Expo-Power, red circles: CRRA. Mean predictions—Boys: $\sigma_u^D = 0.56$ (Expo-Power) vs. 0.46 (CRRA), ICC$^D = 0.43$ for both. Girls: $\sigma_u^D = 0.89$ (Expo-Power) vs. 0.51 (CRRA), ICC$^D = 0.38$ for both. Error bars show 90% confidence intervals.

5.4.4 Effects of IQ and Personality Traits on Guessing Structure

Figure 5 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. $= 0.059$, $p < 0.10$), indicating that emotionally stable boys show more consistent guessing patterns. The baseline correlation is negative ($\rho = -0.15$), meaning boys with persistent traits favoring riskier choices (lower θ^D) tend to guess more often. Extraversion reduces ICC^G (-0.079, $p < 0.10$).

For girls, the correlation structure differs fundamentally: the baseline correlation is positive ($\rho = +0.11$), meaning girls with persistent traits favoring riskier choices tend to deliberate more (guess less). IQ significantly strengthens this positive correlation (coef. = 0.080, $p < 0.001$). These opposite correlation patterns suggest boys and girls employ fundamentally different cognitive strategies when facing difficult choices.

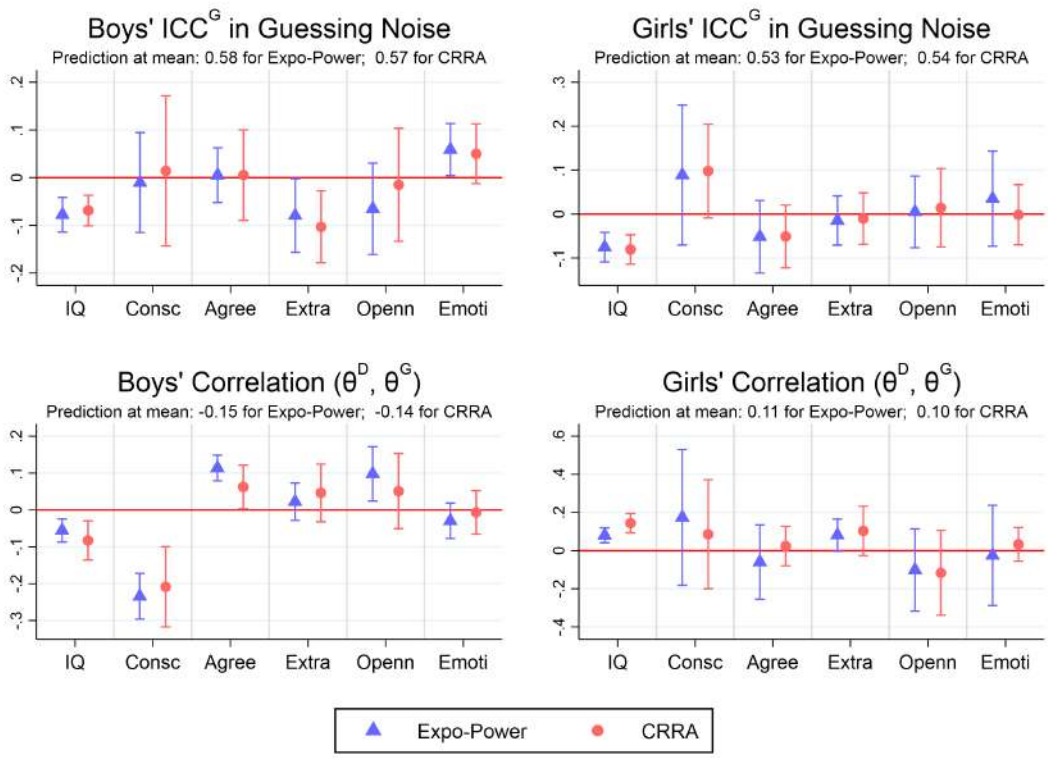


Figure 5: Guessing Structure: ICC and Error Correlation

Notes: 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. $= 0.059$, $p < 0.10$), baseline correlation negative ($\rho = -0.15$). Girls: baseline correlation positive ($\rho = +0.11$), IQ strengthens it (coef. $= 0.080$, $p < 0.001$). Error bars show 90% confidence intervals.

5.4.5 Effects of IQ and Personality Traits on Guessing Probability

Figure 6 presents the effects of IQ and personality traits on overall guessing probability. IQ strongly reduces guessing for both genders (boys: -0.070, $p < 0.001$, girls: -0.100, $p < 0.001$), 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 (-0.058, $p < 0.05$), while no personality trait significantly affects girls' guessing probability. This absence of strong trait effects on overall frequency contrasts with significant effects on ICC^G and correlation (Figure 5), indicating that traits affect the consistency and structure of guessing rather than its frequency. Mean guessing probabilities are 13% for boys and 16% for girls, indicating that strategic disengagement is a meaningful component of children's decision-making.

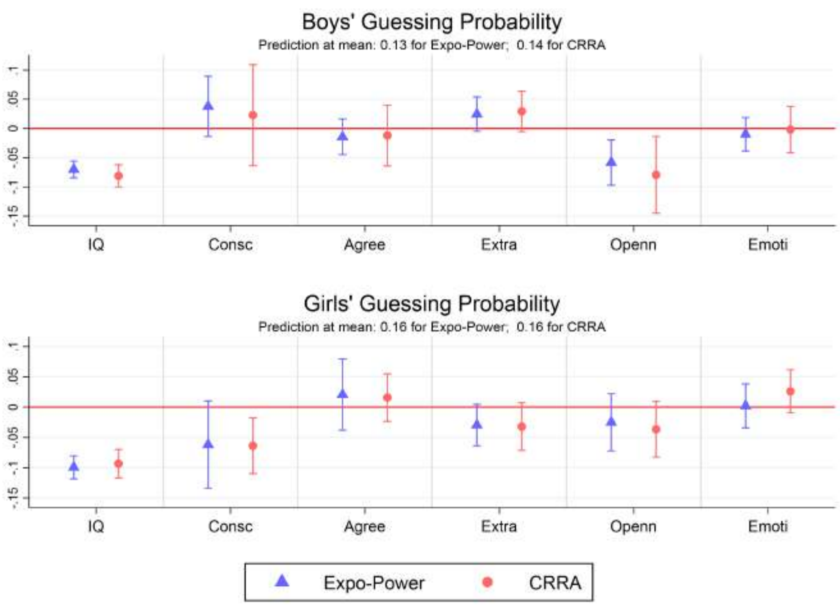


Figure 6: Trait Effects on Guessing Probability

Notes: Effects of IQ and traits on overall guessing probability. Blue squares: Expo-Power, red circles: CRRA. Boys: mean 13%, Girls: mean 16%. 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 5), indicating traits affect consistency and structure rather than frequency. Error bars show 90% confidence intervals.

5.4.6 Joint Significance Tests

In this subsection, we conduct two sets of hypothesis tests. Table 11 presents Wald tests for the joint significance of personality traits on each parameter within gender. Table 12 examines gender differences.

Table 11 presents Wald tests for the joint null hypothesis that all six psychological traits (IQ and five personality traits) have zero effect on parameters for each 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 in the model (Column 8). Psychological traits jointly predict decision quality 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 ICC in deliberation and guessing as well as the correlation between them; for girls, traits primarily affect the correlation structure and guessing probability.

Table 11: Joint Significance Tests: Personality Traits Effects on Model Parameters

	Parameter							
	r (1)	β (2)	σ_{u^D} (3)	ICC^D (4)	ICC^G (5)	Corr(θ^D, θ^G) (6)	Pr(Guess) (7)	All (8)
Boys, Expo-Power								
With ICC and Guessing	0.055	0.051	0.070	0.019	0.003	0.000	0.029	0.000
No ICC, with Guessing	0.162	0.003	0.423				0.134	0.000
With ICC, No Guessing	0.000	0.018	0.012	0.129				0.000
No ICC, No Guessing	0.012	0.018	0.582					0.000
Boys, CRRA								
With ICC and Guessing	0.010		0.061	0.041	0.011	0.001	0.004	0.000
No ICC, with Guessing	0.098		0.852				0.006	0.000
With ICC, No Guessing	0.848		0.203	0.060				0.000
No ICC, No Guessing	0.120		0.552					0.000
Girls, Expo-Power								
With ICC and Guessing	0.421	0.023	0.148	0.388	0.623	0.041	0.053	0.000
No ICC, with Guessing	0.123	0.001	0.000				0.098	0.000
With ICC, No Guessing	0.000	0.000	0.000	0.174				0.000
No ICC, No Guessing	0.000	0.000	0.000					0.001
Girls, CRRA								
With ICC and Guessing	0.220		0.420	0.555	0.643	0.000	0.024	0.000
No ICC, with Guessing	0.297		0.006				0.027	0.000
With ICC, No Guessing	0.076		0.000	0.163				0.000
No ICC, No Guessing	0.267		0.016					0.001

Notes: Each entry is a p-value for the joint null hypothesis that IQ and the five personality traits have no effect on the indicated parameter. Wald tests with 6 degrees of freedom. 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 12 addresses gender differences in the impact of traits. When boys and girls differ in parameters, 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.

To isolate structural from compositional effects, we employ a standard finite approximation to a total differential. We compare specifications based on Expo-Power and CRRA functions. Let $\boldsymbol{b}_l$ denote the vector of coefficients (intercept plus slopes on psychological traits) for gender $l \in \{\text{boys}, \text{girls}\}$, and let $\bar{\boldsymbol{X}}_l$ denote the mean of the psychological traits vector for that gender. For each structural parameter, its predicted value for gender l is at the sample mean $\bar{\boldsymbol{X}}_l'\boldsymbol{b}_l$, where appropriate link functions are applied for bounded parameters. One way to decompose the total gender difference at mean is:

$$\underbrace{\bar{\boldsymbol{X}}_{\text{girls}}'\boldsymbol{b}_{\text{girls}} - \bar{\boldsymbol{X}}_{\text{boys}}'\boldsymbol{b}_{\text{boys}}}_{\text{Total difference } \Delta(\boldsymbol{X}'\boldsymbol{b})} = \underbrace{(\bar{\boldsymbol{X}}_{\text{girls}} - \bar{\boldsymbol{X}}_{\text{boys}})'\boldsymbol{b}_{\text{girls}}}_{\text{Composition effect}} + \underbrace{\bar{\boldsymbol{X}}_{\text{boys}}'(\boldsymbol{b}_{\text{girls}} - \boldsymbol{b}_{\text{boys}})}_{\text{Coefficient effect}} \tag{25}$$

The composition effect captures how much of the gender gap arises because boys and girls have different mean levels of IQ and personality traits. The coefficient effect captures how much arises because the same traits map differently to the parameter across genders—a genuinely structural difference.

For each parameter, Table 12 reports three quantities: (1) p-value for $\Delta(\boldsymbol{X}'\boldsymbol{b})$, which tests whether predicted values differ between genders when each is evaluated at its own covariate mean; (2) p-value for $\Delta\boldsymbol{b}$, which tests the joint hypothesis that all coefficients (intercept and slopes) are equal across genders, isolating the structural component; and (3) % due to $\Delta\boldsymbol{b}$, the share of the total difference attributable to coefficient differences. Values outside $[0, 100]$ indicate that composition and coefficient effects have opposite signs.

Table 12: p-values for Gender Differences Tests in Parameters by Components

	Parameter						
	r	β	σ_{u^D}	ICC^D	ICC^G	Corr(θ^D, θ^G)	Pr(Guess)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Expo-Power, With ICC and Guessing							
p-value for $\Delta(\boldsymbol{X}'\boldsymbol{b})$	0.000	0.847	0.000	0.209	0.211	0.001	0.134
p-value for $\Delta\boldsymbol{b}$	0.000	0.011	0.000	0.094	0.102	0.000	0.041
% Due to $\Delta\boldsymbol{b}$	−9.7	237.1	−17.0	−13.1	21.9	−21.4	39.3
Expo-Power, No ICC, with Guessing							
p-value for $\Delta(\boldsymbol{X}'\boldsymbol{b})$	0.348	0.001	0.546				0.246
p-value for $\Delta\boldsymbol{b}$	0.000	0.000	0.000				0.009
% Due to $\Delta\boldsymbol{b}$	−16.2	3.5	2.4				48.9
Expo-Power, With ICC, No Guessing							
p-value for $\Delta(\boldsymbol{X}'\boldsymbol{b})$	0.002	0.298	0.003	0.094			
p-value for $\Delta\boldsymbol{b}$	0.000	0.000	0.000	0.161			
% Due to $\Delta\boldsymbol{b}$	−9.2	25.6	−12.4	11.7			
Expo-Power, No ICC, No Guessing							
p-value for $\Delta(\boldsymbol{X}'\boldsymbol{b})$	0.345	0.055	0.704				
p-value for $\Delta\boldsymbol{b}$	0.000	0.000	0.000				
% Due to $\Delta\boldsymbol{b}$	−11.6	0.3	−37.2				
CRRA, With ICC and Guessing							
p-value for $\Delta(\boldsymbol{X}'\boldsymbol{b})$	0.931		0.163	0.063	0.394	0.000	0.198
p-value for $\Delta\boldsymbol{b}$	0.009		0.004	0.227	0.216	0.000	0.070
% Due to $\Delta\boldsymbol{b}$	−773.9		−40.5	−7.4	11.2	−21.0	34.4
CRRA, No ICC, with Guessing							
p-value for $\Delta(\boldsymbol{X}'\boldsymbol{b})$	0.000		0.012				0.217
p-value for $\Delta\boldsymbol{b}$	0.000		0.131				0.444
% Due to $\Delta\boldsymbol{b}$	2.6		−2.5				25.3
CRRA, With ICC, No Guessing							
p-value for $\Delta(\boldsymbol{X}'\boldsymbol{b})$	0.363		0.278	0.110			
p-value for $\Delta\boldsymbol{b}$	0.508		0.224	0.167			
% Due to $\Delta\boldsymbol{b}$	29.0		−10.3	13.8			
CRRA, No ICC, No Guessing							
p-value for $\Delta(\boldsymbol{X}'\boldsymbol{b})$	0.000		0.000				
p-value for $\Delta\boldsymbol{b}$	0.000		0.003				
% Due to $\Delta\boldsymbol{b}$	4.0		3.1				

Notes: For each parameter and specification, we report three quantities. "p-value for $\Delta(\boldsymbol{X}'\boldsymbol{b})$" tests whether predicted values differ between boys and girls when each gender is evaluated at its own mean of IQ and personality traits. "p-value for $\Delta\boldsymbol{b}$" tests the joint null hypothesis that all coefficients (intercept and slopes on IQ and traits) are equal across genders. "% Due to $\Delta\boldsymbol{b}$" reports the percentage of the total gender difference attributable to coefficient differences rather than differences in mean covariates, based on the standard decomposition. Percentages outside $[0, 100]$ indicate that composition and coefficient effects have opposite signs. Empty cells indicate parameters not estimated in that specification.

The decomposition reveals that gender differences in risk preferences reflect both compositional and structural sources, often operating in opposite directions. Compositional differences refer to gender gaps arising because boys and girls have different mean levels of traits (e.g., girls score higher on Conscientiousness). Structural differences refer to gaps arising because the same trait maps differently to outcomes by gender (e.g., Conscientiousness increases r for boys but not girls). The decomposition separates these two sources: if the structural share is near 100%, the gap reflects different effects of traits; if near 0%, the gap reflects different trait endowments evaluated at common returns.

Under the full Expo-Power specification, both the total gap and structural difference in baseline curvature r are highly significant ($p = 0.000$ for both). However, the decomposition shows that only -9.7% of the total gap is due to coefficient differences. The negative sign indicates that composition and structure work in opposite directions: using girls' structural relationships, their trait distribution would predict *lower* r than boys (composition accounts for 110% of the gap). The coefficient differences partially offset this, as they work in girls' favor. The net result is that boys have higher r. For wealth-sensitivity β, the pattern reverses: predicted values do not differ ($p = 0.847$), but coefficients do ($p = 0.011$), with 237% of the (near-zero) total gap due to structural differences. This indicates that boys and girls arrive at similar average β through different trait channels—a case of offsetting composition and structure effects.

Under CRRA with ICC and guessing, the total gender difference in r vanishes ($p = 0.931$), yet the structural test remains statistically significant ($p = 0.009$). The decomposition shows -773.9% due to coefficients, indicating that large structural differences are almost perfectly offset by compositional differences. This confirms that

CRRA's functional form restrictions mask genuine gender differences: the underlying structural relationships differ substantially, but the restrictive parameterization compresses predicted values toward equality.

The persistent components ICC^D and ICC^G show no significant gender differences in either total gap or structure ($p > 0.09$ for all tests), suggesting that the proportion of decision variance attributable to stable traits is similar across genders. Transitory noise σ_{u^D} differs significantly in both total gap ($p = 0.000$) and structure ($p = 0.000$), with -17% due to coefficients—indicating that girls' higher noise is primarily compositional (driven by trait differences) rather than structural.

The correlation $\rho(\theta^D, \theta^G)$ shows highly significant differences in both tests ($p = 0.001$ for total, $p = 0.000$ for structure), with -21.4% due to coefficients. This confirms that the opposite-signed correlations—negative for boys ($\rho = -0.15$) and positive for girls ($\rho = +0.11$)—reflect genuinely different cognitive architectures. The structural difference is robust: even holding trait distributions constant, boys and girls exhibit fundamentally different relationships between deliberation precision and guessing propensity.

The decomposition clarifies how to interpret the trait-by-trait coefficient plots presented earlier. We observe, for instance, that Conscientiousness significantly increases r for boys but not for girls. This is not a mechanical artifact of boys and girls having different mean levels of Conscientiousness. The $\Delta\boldsymbol{b}$ test confirms that these coefficient differences are statistically significant: the same trait operates through different channels depending on gender. Likewise, when Extraversion predicts r for girls but not boys, this reflects a true structural difference in how personality maps to risk preferences. Our decomposition shows that these structural differences are not merely compositional consequences of girls scoring higher on Conscientiousness or Agreeableness. Rather, the economic consequences of personality traits—how much a one-unit increase in a trait shifts a given parameter—differ fundamentally by gender. This supports the broader conclusion that boys and girls exhibit distinct cognitive architectures in economic decision-making.

5.4.7 Which Traits Matter and Through Which Channels?

Table 13 summarizes the significant trait effects by gender and channel. The findings demonstrate that 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 decision quality, with particularly strong effects on decision precision and guessing behavior. For girls, traits also affect both channels, but through different specific traits—Extraversion rather than Conscientiousness drives preference effects. This gender difference in which traits matter, confirmed by the significant $\Delta \boldsymbol{b}$ tests in Table 12, supports the conclusion that boys and girls exhibit distinct mappings from personality to economic behavior.

Table 13: Summary of Significant Trait Effects

Gender	Trait	Channel	Effect
Boys	Conscientiousness	Risk aversion (r)	Increases (+0.104**)
	Conscientiousness	Decision precision (σ_{u^D})	Improves ($-$0.201***)
	IQ	Decision precision (σ_{u^D})	Improves ($-$0.103*)
	IQ	Guessing probability	Reduces ($-$0.070***)
	Emotional Stability	Guessing consistency (ICC^G)	Increases (+0.059*)
	Openness	Wealth-sensitivity (β)	Reduces ($-$0.043**)
Girls	Extraversion	Risk aversion (r)	Increases (+0.090*)
	Extraversion	Wealth-sensitivity (β)	Increases (+0.017*)
	IQ	Guessing probability	Reduces ($-$0.100***)
	IQ	Correlation (ρ)	Strengthens (+0.080***)

Notes: Significant effects at *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$ from full Expo-Power model. Effects reported are coefficient estimates. Signs indicate direction of effect.

5.5 Alternative Guessing Specifications

We compare three specifications of equation (9): (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 9, 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):

$$\eta(\boldsymbol{X}_i) = \sum_{k=1}^{K} \eta_k \cdot \mathrm{X}_{ki}, \tag{26}$$

where η_k is the parameter of interest.

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

$$\eta(\sigma^2_{u^D}(\boldsymbol{X}_i)) = \eta_0 + \eta_\sigma \cdot \sigma^D_{u,i}, \tag{27}$$

with $\sigma^2_{u^D}(\boldsymbol{X}_i)$ being the variance of the deliberation noise defined in Equation 24.

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

$$\eta(\sigma^2_{u^D}(\boldsymbol{X}_i), \boldsymbol{X}_i) = \eta_\sigma \cdot \sigma^D_{u,i} + \sum_{k=1}^{K} \eta_k \cdot \mathrm{X}_{ki}. \tag{28}$$

In all three models, we include ICC, guessing, and the effects of IQ and personality traits, consistent with our full baseline specification. Model 3 achieves marginally better AIC. Tables and figures in Appendix I present detailed parameter estimates for the three alternative specifications. The models yield similar marginal effects of IQ and personality traits on most parameters, except for the guessing probability.

Table 14: Model Comparison: Alternative Guessing Specifications

Panel A: Boys	Model 1 Only Traits	Model 2 Only Cognitive Difficulty	Model 3 Joint
Loglikelihood	−32755.039	−32766.759	−32752.633
k (parameters)	49	44	50
AIC	65608.08	65619.60	65605.27

Panel B: Girls	Model 1 Only Traits	Model 2 Only Cognitive Difficulty	Model 3 Joint
Loglikelihood	−33232.537	−33238.099	−33230.045
k (parameters)	49	44	50
AIC	66563.07	66564.20	66560.09

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. Model 2 specifies that deliberation noise directly influences guessing, with traits affecting guessing only indirectly through their impact on deliberation noise. Model 3 includes both channels. Lower AIC indicates better model fit.

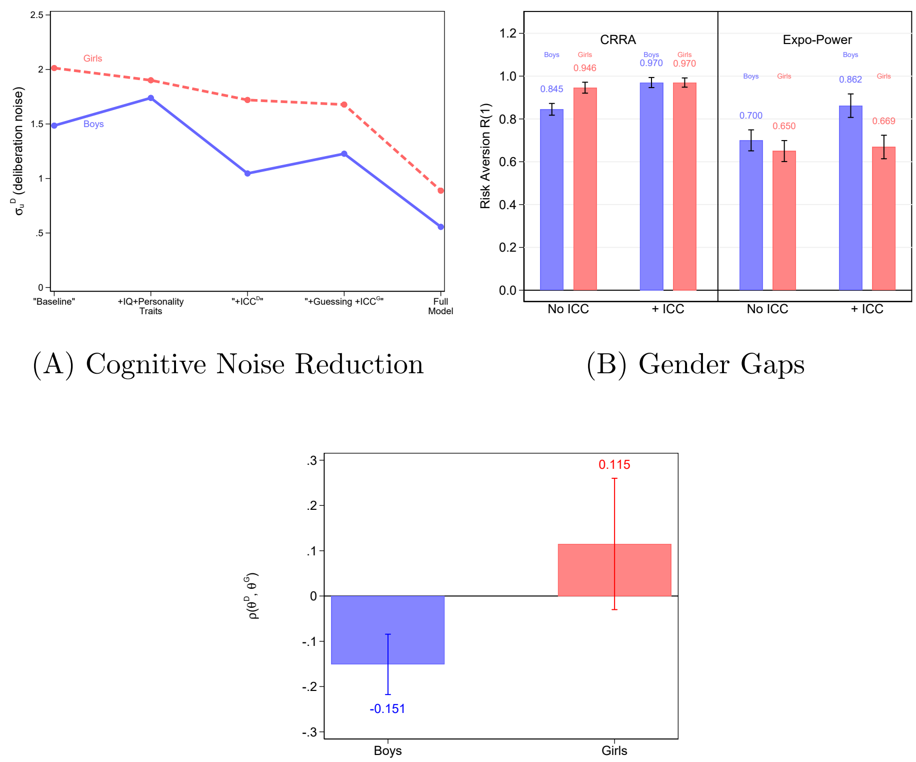
For girls, the marginal effect of deliberation noise σ_{u^D} on guessing probability is large and highly significant when traits are restricted to operate only through noise variances (Model 2). Under Expo-Power, a one-unit increase in σ_{u^D} raises guessing probability by 4.81 percentage points ($p < 0.01$). Under CRRA, the effect is even larger at 14.20 percentage points ($p < 0.01$). However, when direct trait effects are added (Model 3), these coefficients become small and insignificant (-0.85 and -3.82, respectively), indicating that the apparent noise channel largely proxies for underlying personality differences that affect guessing directly.

For boys, the deliberation noise channel is weak throughout. Model 2 yields positive but insignificant effects (0.93 under Expo-Power, 2.50 under CRRA), while Model 3 produces near-zero or negative coefficients (-0.41 and -3.33, respectively). These patterns suggest a fundamental gender asymmetry: girls are more likely to disengage when facing imprecise deliberation signals, whereas boys' guessing behavior is largely unresponsive to decision difficulty. The finding that direct trait effects absorb the noise channel for both genders—but especially for girls—indicates that cognitive and personality characteristics influence guessing through multiple pathways, not solely through their impact on deliberation precision.

5.6 Key Findings and the Gender Gap Reversal

Figure 7 consolidates our main findings. Our analysis overturns the conventional finding that girls are more risk-averse than boys. Standard CRRA without accounting for decision quality replicates the traditional pattern (girls 12% higher). However, controlling for decision quality through ICC eliminates the gap entirely (both

$R(x, w) = 0.97$). Expo-Power then reveals the opposite: boys are 29% more risk-averse than girls—a 34 percentage-point reversal. ICC analysis reveals that 40–45% of apparent measurement error reflects stable 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. $\rho = +0.11$): boys with unobserved persistent traits favoring riskier choices tend to guess more often, whereas girls with such traits are more likely to deliberate.



(A) Cognitive Noise Reduction

(B) Gender Gaps

(C) Deliberation-Guessing Correlation

Figure 7: Cognitive Noise, Gender Gaps, and Decision Processes

Notes: Panel A sequentially decomposes decision variance into persistent differences (40–45%) versus transitory shocks. Panel B shows how measurement and functional forms jointly determine gender gaps: CRRA without ICC replicates conventional findings that girls are more risk-averse than boys. CRRA with ICC narrows this gender gap. Expo-Power with ICC reverses it, showing boys are 29% more risk-averse than girls. Panel C shows the estimated correlation between the persistent component of errors in deliberation and guessing for boys and girls. Error bars represent 90% confidence intervals.

Component interactions are documented in Appendix O. For boys' Expo-Power, components estimated separately yield 2,377.0 log-likelihood points improvement, yet the full model improves by 3,259.7 points—an 882.7-point superadditivity (37% beyond additive). For girls, superadditivity equals 538.4 points (25% beyond the

2,164.4-point sum). This positive interaction confirms that preference recovery requires modeling preferences, precision, and engagement simultaneously.

5.7 Heterogeneity in Risk Preferences

Figure 8 displays histograms of predicted risk aversion $R(x, w) = r(\text{IQ}, \text{traits}) + \beta(\text{IQ}, \text{traits}) \cdot (x + w)^{1-r(\text{IQ},\text{traits})}$ evaluated at $x = 0$ and $w = 1$ (i.e., baseline wealth before any experimental payoff), using individual-specific covariates and estimated parameters from the full Expo-Power model.

The distributions reveal substantial within-gender heterogeneity and a striking between-gender difference. Boys exhibit mean risk aversion of 0.86 (SD = 0.08) with roughly symmetric distribution. Girls exhibit a substantially lower mean risk aversion of 0.67 (SD = 0.08) with modest right skew. Within-gender variation is comparable across genders (identical standard deviations), reflecting heterogeneity in IQ and personality traits.

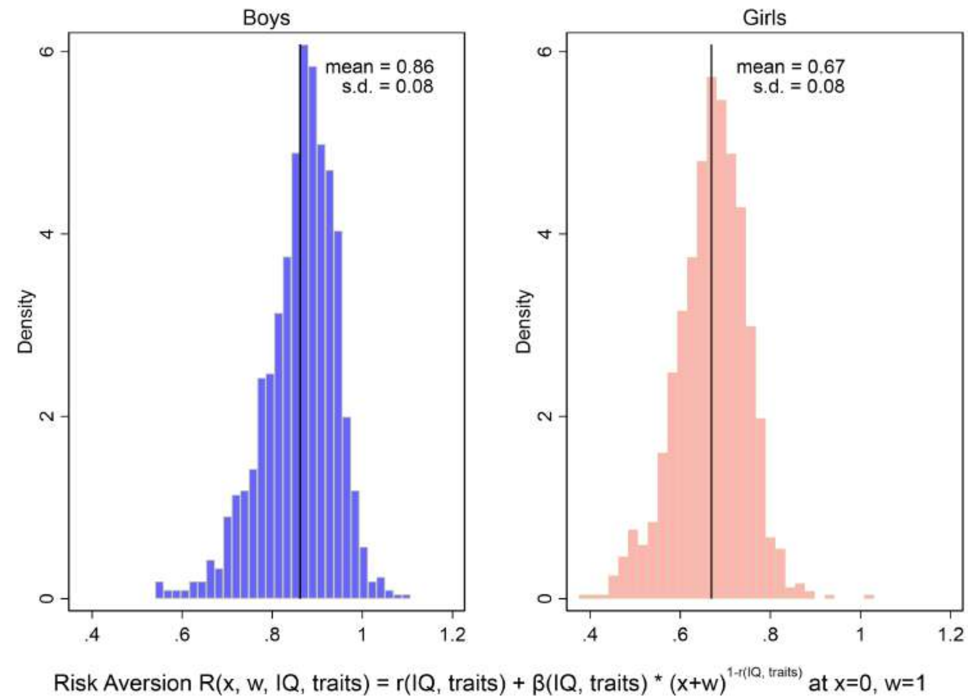


Figure 8: Distribution of Individual-Level Risk Aversion Estimates

Notes: Histograms of predicted risk aversion $R(x, w) = r + \beta(x + w)^{1-r}$ evaluated at $x = 0$ and $w = 1$ using individual traits from the full Expo-Power specification. Boys (blue, $N = 1{,}115$): mean $= 0.86$, SD $= 0.08$. Girls (red, $N = 1{,}089$): mean $= 0.67$, SD $= 0.08$. Vertical lines indicate means.

5.8 Robustness

Our main findings are robust to alternative specifications of socioeconomic controls, fallback consumption levels, and reference wealth heterogeneity. We summarize each robustness check below. Full results appear in Appendices L–N.

5.8.1 The Role of Background Variables

A large 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. In this section we explore the role of background variables on shaping child risk preferences. We find that background affects psychological traits, but conditioning on those traits there is no additional role for background variables. When added to our models when traits are included, they have no effect on any component of our models, both for CRRA and Expo-Power.

Tables C4 and C5 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. Figures L1–L8 in Appendix L present results for our preferred Expo-Power specifications, while Figures L9–L14 present results for CRRA. A striking finding is near-identical predictions across all three models under both functional forms.

Under Expo-Power, overall risk aversion $R(x, w)$ evaluated at $x = 0$ and $w = 1$ shows remarkable stability across the three specifications of covariates. Boys have a risk aversion of $R(x, w) = 0.85$–0.86 across the three specifications, while girls have $R(x, w) = 0.67$ across all specifications. Model 1 (background only) yields $r = 0.89$ and $\beta = -0.04$, whereas Models 2 and 3 yield $r = 0.76$–0.77 and $\beta = 0.09$. The baseline curvature and wealth-sensitivity move in opposite directions. Overall risk aversion remains constant. 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 role between level and slope components. For girls the decomposition is robust to covariate choice.

Under CRRA, risk aversion results are similar. Boys exhibit $r = 0.97$–0.98 across all three specifications, while girls show $r = 0.97$ uniformly. The same pattern holds for decision quality parameters. Under Expo-Power, transitory noise σ_u^D shows predictions of 0.54–0.56 for boys and 0.85–0.89 for girls across models; under CRRA, predictions are 0.45–0.46 for boys and 0.50–0.52 for girls. ICC^D remains at 0.42–0.43 for boys and 0.38–0.39 for girls under Expo-Power, and 0.41–0.43 for boys and 0.37–0.39 for girls under CRRA.

For ICC^G, boys show a modest 10% reduction when adding IQ and traits under both specifications (0.64 in Model 1 to 0.58–0.59 in Models 2 and 3 under Expo-Power; 0.63 to 0.57 under CRRA), reflecting the genuine predictive power of personality traits—particularly Emotional Stability—for persistent guessing patterns. Girls' ICC^G shows a smaller reduction from 0.57–0.58 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 perhaps the most striking evidence. The gender difference in correlation patterns—negative for boys ($\rho = -0.15$ to -0.19) and positive for girls ($\rho = 0.06$ to 0.11)—persists across all three specifications under Expo-Power. The same pattern holds under CRRA: boys exhibit $\rho = -0.14$ to -0.18, while girls exhibit $\rho = 0.09$ to 0.10. 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 0.11–0.13 for boys and 0.15–0.16 for girls across models under Expo-Power, and 0.12–0.14 for boys and 0.16 for girls under CRRA, stable within 2 percentage points.

Our analysis reveals that background variables have little direct effect on economic preferences once psychological traits are accounted for. 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 decision quality indirectly—by shaping cognitive ability and personality—rather than through direct channels.

5.8.2 Sensitivity to Choice of Fallback Consumption Level

Our main analyses assume a fallback consumption level of $w_0 = 1$, representing consumption when gaining nothing from the experimental game. Appendix M examines sensitivity to this normalization by re-estimating all models with $w_0 \in \{0.2, 0.5, 1, 2, 5\}$ for both Expo-Power and CRRA specifications.

The level of risk aversion naturally varies with assumed fallback consumption. In the Expo-Power specification, boys' risk aversion $R(x, w)$ ranges from 0.75 at $w_0 = 0.2$ to 1.19 at $w_0 = 5$, while girls' ranges from 0.60 to 1.05 across the same values. The CRRA specification shows similar patterns: boys' r ranges from 0.91 to 1.16, and girls' from 0.91 to 1.16. These shifts are mechanical consequences of evaluating curvature at different consumption levels and do not affect our substantive conclusions.

The critical finding is that structural parameters governing decision quality remain remarkably stable. ICC^D is essentially invariant to choice of fallback consumption level. For boys, it ranges from 0.42 to 0.44 across all w_0 values in both specifications. For girls, it remains at 0.37–0.38. This stability confirms that the persistent component of deliberation precision—approximately 40% of total variance—is a robust structural feature independent of normalization choices.

ICC^G shows similar invariance: boys maintain ICC$^G \approx 0.57$–0.58 across all fallback levels, while girls show ICC$^G \approx 0.52$–0.54. The correlation between deliberation and guessing random effects is equally robust: boys consistently show $\rho \approx -0.14$ to -0.17, while girls show $\rho \approx 0.09$ to 0.13. The gender difference in correlation sign persists across all 10 specification-by-fallback combinations (5 fallback levels $\times$ 2 functional forms), confirming that this reflects genuine behavioral differences rather than estimation artifacts.

Guessing probability is virtually identical across choice of fallback levels: boys show 0.13–0.14 regardless of w_0, while girls show 0.16–0.17. The coefficient patterns for IQ and personality traits similarly maintain their qualitative features—Conscientiousness remains the dominant predictor for boys, Extraversion for girls—though point estimates shift somewhat with fallback consumption.

Most importantly, our main finding regarding gender differences in risk attitudes remains unchanged across all specifications: CRRA without ICC reproduces the conventional finding that girls are more risk-averse than boys. CRRA with ICC narrows this gap, and Expo-Power with ICC reverses it. The robustness of this pattern across fallback consumption levels strengthens confidence that the gender gap reversal reflects genuine preference differences rather than normalization artifacts.

5.8.3 Reference Wealth and Socioeconomic Background

A related concern is whether children from different socioeconomic backgrounds develop systematically different reference points based on their lived experiences. Appendix N 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.

Figures N1–N7 present results for the Expo-Power 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 0.86 to 0.99 for boys and from 0.67 to 0.60 for girls—but 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 quality 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 0.13 for boys and 0.15–0.16 for girls regardless of whether reference wealth is fixed or estimated.

An interesting finding emerges from the estimated reference wealth function itself.

For boys, predicted fallback consumption at mean background variables is $w = 2.04$, suggesting boys may integrate the experimental stakes within a broader consumption choice. For girls, predicted fallback consumption is essentially zero ($w = 0.00$), indicating that girls may evaluate experimental gambles in isolation from background consumption. This gender difference in reference point formation—boys integrating stakes with background wealth, girls evaluating stakes narrowly—could contribute to observed differences in risk-taking behavior, though the mechanism warrants further investigation.

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 once decision quality is properly modeled.

6 Understanding Variability in Choices

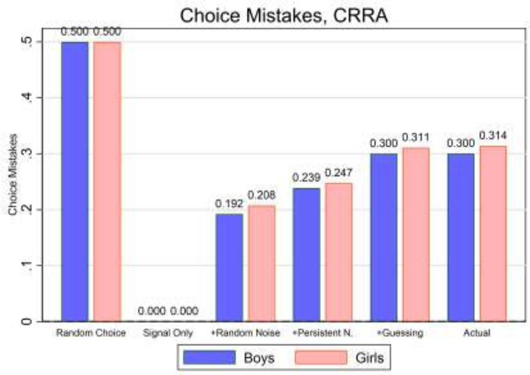
Having established how cognitive skills and personality traits affect risk preferences, deliberation noise, and guessing, we now quantify the relative importance of these three channels for explaining observed choice mistakes in the full developed model. This analysis operates under the assumption that the previous sections recovers true parameters and that we have correctly identified the components in Figure 1. We assess how much variation in choices reflects true preference differences versus decision imprecision, and whether imprecision stems from stable unobserved characteristics in deliberation and guessing, or transitory shocks.

6.1 Sequential Decomposition

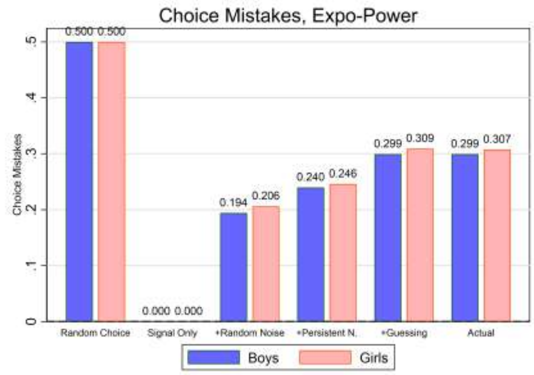
Figures 9a and 9b present sequential decompositions of choice mistakes under CRRA and Expo-Power specifications. Each figure displays six benchmarks representing progressive inclusion of model components: (1) the theoretical mistake rate of 0.50 if children flipped coins, (2) the theoretical rate of zero with perfect decision-making, (3) the simulated rate with transitory random deliberation noise (u_{ij}^D) only, (4) the rate adding persistent individual noise (θ_i^D), (5) the rate with all three components including guessing, and (6) the estimated mistake rate 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 21–22% to the total mistake rate, persistent individual noise adds 4–5%, and guessing contributes 4–5%. The close alignment between full model predictions and actual observed choices validates the model's ability to capture decision error structure.

The Expo-Power and CRRA models yield remarkably similar decompositions, with total mistake rates differing by less than one percentage point (30.0% vs. 30.0% for boys, 30.7% vs. 31.1% for girls). This stability suggests the decomposition is robust to functional form assumptions.



(a) CRRA Specification



(b) Expo-Power Specification

Figure 9: Sequential Decomposition of Choice Mistakes

Notes: Sequential decomposition of choice mistakes under CRRA (left) and Expo-Power (right) specifications. Blue bars represent boys, red bars represent girls. "Random Choice" reflects the theoretical 50% mistake rate. "Signal Only" shows a 0% rate under perfect decision-making. "+ Random Noise" introduces transitory noise (u_{ij}^D). "+ Persistent N." adds persistent noise (θ_i^D) on top. "+ Guessing" activates all remaining components associated with guessing: u_{ij}^G, θ_i^G, and the correlation $\rho(\theta_i^D, \theta_i^G)$. "Actual" reflects the empirical mistake rate, calculated as the difference between observed choices and model-predicted choices in the absence of error or guessing.

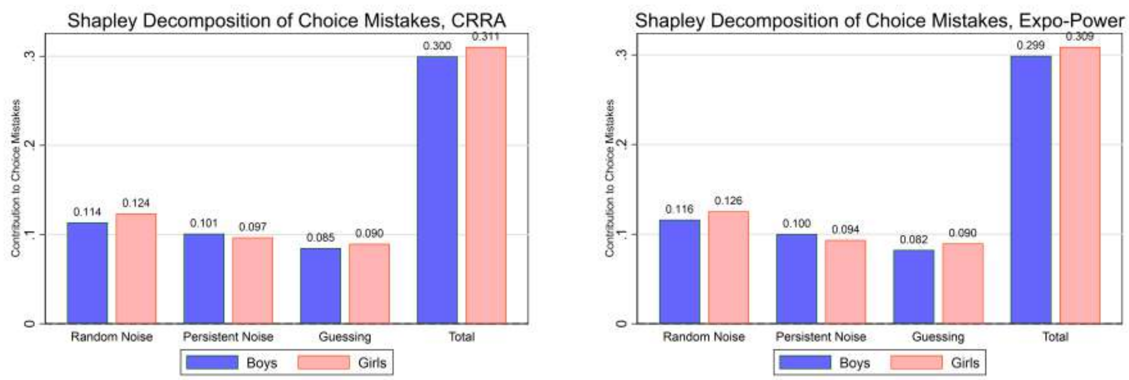
6.2 Shapley Decomposition

Figures 10a and 10b present Shapley value decompositions, which compute each component's average marginal contribution across all possible orderings, addressing the path dependence problem of any sequential decomposition.

Random transitory noise (u_{ij}^D) accounts for approximately 11–13% of total choice mistakes across both genders and specifications. Persistent individual noise (θ_i^D) contributes roughly 9–10%. Guessing behavior accounts for approximately 9%. Total mistake rates of approximately 30% are consistent with the sequential decomposition and observed data.

All three error sources make substantial and roughly comparable contributions. Random noise is the largest single contributor, but persistent noise and guessing each account for a substantial amount6 of variation. This contradicts models that attribute all choice inconsistency to pure randomness.

Gender differences are modest. Girls exhibit slightly higher contributions from random noise (12.6%) compared to boys (11.6% under Expo-Power), while persistent noise and guessing show similar contributions.



(a) CRRA Specification

(b) Expo-Power Specification

Figure 10: Shapley Decomposition of Choice Mistakes

Notes: Shapley value decomposition of choice mistakes under CRRA (left) and Expo-Power (right) specifications. Blue bars: boys, red bars: girls. Each component's contribution computed as 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 disengagement. "Total" shows the sum.

6.3 Implications for Preference Measurement

These decompositions carry important methodological implications. Traditional approaches that attribute all choice inconsistency to shocks to preferences systematically overestimate the role of noise and ignore substantial contributions of persistent individual differences and guessing behavior. Our findings suggest approximately 70% of observed mistakes stem from random transitory noise, with persistent individual differences and guessing each accounting for roughly 15%.

This evidence has practical consequences. First, researchers should account for persistent heterogeneity in decision quality when estimating preference parameters. Failure to do so conflates true preference variation with variation in cognitive precision. Second, tasks inducing high guessing rates—through complexity, length, or weak incentives—distort preference estimates. Third, the substantial role of transitory noise suggests within-subject repeated measures can improve precision by averaging over task-specific noise, though persistent components remain.

The robustness of these decompositions across CRRA and Expo-Power specifications demonstrates that the underlying structure of decision errors is not an artifact 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 Conclusion

This paper addresses a fundamental challenge in preference measurement: when decision quality varies systematically across individuals, standard approaches confound true preference heterogeneity with differences in the precision with which those preferences are expressed. We demonstrate this problem in the context of children's risk preferences, where conventional methods have consistently found girls more risk-averse than boys across five decades of research. Our findings overturn this consensus, revealing that the apparent gender gap reflects error and functional form restrictions rather than true preference differences.

We establish two core results. First, properly accounting for decision quality reverses the documented gender gap. Standard CRRA specifications replicate the conventional finding: girls appear more risk-averse than boys. However, this gap disappears entirely when we model persistent individual differences in decision precision through ICC structure. The flexible Expo-Power specification then reveals boys are actually substantially more risk-averse than girls—a complete reversal of the conventional pattern.

Second, we decompose total decision variance into persistent noise versus transitory noise, finding that approximately 40% of apparent deliberation error actually reflects stable individual differences in decision quality. Baseline models that treat all inconsistency as random noise severely overestimate cognitive imprecision. Once we separate these components, girls exhibit higher transitory noise than boys even after controlling for persistent noise, indicating gender differences operate through both stable decision quality and moment-to-moment precision.

Our structural framework advances preference measurement by jointly modeling preferences through flexible utility functions, deliberation precision through ICC-structured noise, and strategic engagement through explicit guessing behavior. Components interact positively rather than additively. The full model improves fit substantially beyond the sum of individual components, confirming that proper identification of preference requires simultaneous treatment of preferences, deliberation and guessing.

The random effect ICC framework provides powerful leverage for separating persistent heterogeneity from transitory noise. The substantial ICC values we estimate indicate that treating choice inconsistency as independent deliberation errors fundamentally mischaracterizes how children make decisions. Allowing the correlation between deliberation and guessing errors proves both empirically essential and theoretically revealing.

IQ and personality traits operate through multiple channels that vary by gender. For boys, traits predominantly affect deliberation precision, guessing, and error correlations rather than directly affecting preferences. For girls, traits influence both preference parameters and decision quality.

These findings likely extend beyond childhood preference measurement. Any population 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 decision quality varies systematically.

The gender differences we document raise questions about developmental origins and long-term consequences. Our longitudinal data enable future analysis of how these patterns emerge and evolve as children mature, potentially revealing critical periods when interventions might be most effective. Understanding whether personality affects behavior primarily through preferences or decision processes has direct implications for intervention design: targeting noncognitive skill development may improve choice quality without shifting underlying preferences, suggesting distinct pathways for educational and developmental policy.

More fundamentally, our results challenge the conventional analysis of error in choices. Economists typically assume that utility maximization is subject to i.i.d. errors. We demonstrate this assumption fails dramatically. Deliberation errors exhibit substantial persistence, correlate with guessing in gender-specific patterns, and respond systematically to cognitive and personality traits. Our variance decompositions show that persistent noise, transitory noise, and strategic disengagement each contribute substantially to total choice variation.

The structural approach we develop provides a template for settings where decision quality 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.

References

Alan, S. and Ertac, S. (2018). Fostering patience in the classroom: Results from randomized educational intervention. *Journal of Political Economy*, 126(5):1865–1911.

Almlund, M., Duckworth, A. L., Heckman, J. J., and Kautz, T. (2011). Personality psychology and economics. In Hanushek, E. A., Machin, S., and Wößmann, L., editors, *Handbook of the Economics of Education*, volume 4, chapter 1, pages 1–181. Elsevier B. V., Amsterdam.

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

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

Andersson, O., Holm, H. J., Tyran, J.-R., and Wengström, E. (2020). Robust inference in risk elicitation tasks. *Journal of Risk and Uncertainty*, 60(1):1–26.

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

Arrow, K. J. (1965). *Aspects of the Theory of Risk-Bearing*. Yrjö Jahnsson Foundation, Helsinki.

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

Belzil, C. and Jagelka, T. (2024). Separating preferences from endogenous effort and cognitive noise in observed decisions. CREST Working Papers Series No. 2024-13.

Benjamin, D. J., Bismark, M., and Kim, S. (2018). The effect of information and incentives on belief formation. *American Economic Review*, 108(3):637–671.

Borghans, L., Duckworth, A. L., Heckman, J. J., and ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of Human Resources*, 43(4):972–1059.

Burks, S. V., Carpenter, J. P., Goette, L., and Rustichini, A. (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proceedings of the National Academy of Sciences*, 106(19):7745–7750.

Caplin, A. (2025). *An Introduction to Cognitive Economics: The Science of Mistakes*. MIT Press, Cambridge, MA.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics*, 25(3):573–578.

Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2):448–474.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9(3):522–550.

Domencich, T. A. and McFadden, D. L. (1975). *Urban Travel Demand: A Behavioral Analysis*. Contributions to Economic Analysis. North-Holland Publishing Company, Amsterdam.

Enke, B., Gneezy, U., Hall, B., Martin, D. C., Nelidov, V., Offerman, T., and van de Ven, J. (2023). Cognitive biases: Mistakes or missing stakes? *The Review of Economics and Statistics*, 105(4):818–832.

Feng, S., Han, Y., Heckman, J. J., and Kautz, T. (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.

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

Harless, D. W. and Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica*, 62(6):1251–1289.

Harrison, G. W., Lau, M. I., and Rutström, E. E. (2007). Estimating risk attitudes in Denmark: A field experiment. *Scandinavian Journal of Economics*, 109(2):341–368.

Heckman, J. J., Liu, B., Lu, M., and Zhou, J. (2025). Treatment effects and the measurement of skills in a prototypical home visiting program. *Journal of Labor Economics*. Forthcoming. NBER Working Paper No. 27356.

Heckman, J. J., Pinto, R., and Savelyev, P. A. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, 103(6):2052–2086.

Hey, J. D. and Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica*, 62(6):1291–1326.

Hill, M. S., Yeung, W.-J. J., and Duncan, G. J. (2001). Childhood family structure and young adult behaviors. *Journal of Population Economics*, 14(2):271–299.

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

Jagelka, T. (2024). Are economists' preferences psychologists' personality traits? a structural approach. *Journal of Political Economy*, 132(3):910–970.

Kautz, T. and Zanoni, W. (2019). 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, G. (2000). Emotions in economic theory and economic behavior. *American Economic Review*, 90(2):426–432.

Malmendier, U. and Nagel, S. (2011). Depression babies: Do macroeconomic experiences affect risk taking? *The Quarterly Journal of Economics*, 126(1):373–416.

Malmendier, U. and Wachter, J. A. (2024). Memory of past experiences and economic decisions. In Kahana, M. J. and Wagner, A. D., editors, *The Oxford Handbook of Human Memory*. Oxford University Press.

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica*, 32(1-2):122–136.

Rabin, M. (1998). Psychology and economics. *Journal of Economic Literature*, 36(1):11–46.

Saha, A. (1993). Expo-power utility: A ‘flexible’ form for absolute and relative risk aversion. *American Journal of Agricultural Economics*, 75(4):905–913.

Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398):605–610.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690.

Soto, C. J., John, O. P., Gosling, S. D., and Potter, J. (2011). Age differences in personality traits from 10 to 65: Big five domains and facets in a large cross-sectional sample. *Journal of Personality and Social Psychology*, 100(2):330–348. Published online 2008, in print 2011.

Sutter, M., Zoller, C., and Glätzle-Rützler, D. (2019). Economic behavior of children and adolescents–a first survey of experimental economics results. *European Economic Review*, 111:98–121.

Tanaka, Y. and Munro, A. (2014). Regional variation in risk and time preferences: Evidence from a large-scale field experiment in rural uganda. *Journal of African Economies*, 23(1):151–187.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34(4):273–286.

Tymula, A., Rosenberg Belmaker, L. A., Roy, A. K., Ruderman, L., Manson, K., Glimcher, P. W., and Levy, I. (2012). Adolescents' risk-taking behavior is driven by tolerance to ambiguity. *Proceedings of the National Academy of Sciences*, 109(42):17135–17140.

von Gaudecker, H.-M., van Soest, A., and Wengström, E. (2011). Heterogeneity in risky choice behavior in a broad population. *American Economic Review*, 101(2):664–694.

Woodford, M. (2020). Modeling imprecision in perception, valuation, and choice. *Annual Review of Economics*, 12(1):579–601.

Xie, D. (2000). Power risk aversion utility functions. *Annals of Economics and Finance*, 1(2):265–282. Working Paper version: CEMA Working Paper No. 22 (Oct 1999).

Appendix

A Appendix A: 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 in the Expo-Power framework with full ICC structure and guessing: Model 1 uses teacher-reported traits (baseline), Model 2 uses self-reported traits. The comparison reveals both convergence and divergence: structural parameters (noise, ICC, guessing) remain stable within 5–10%, but 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.

A.1 Expo-Power Specification

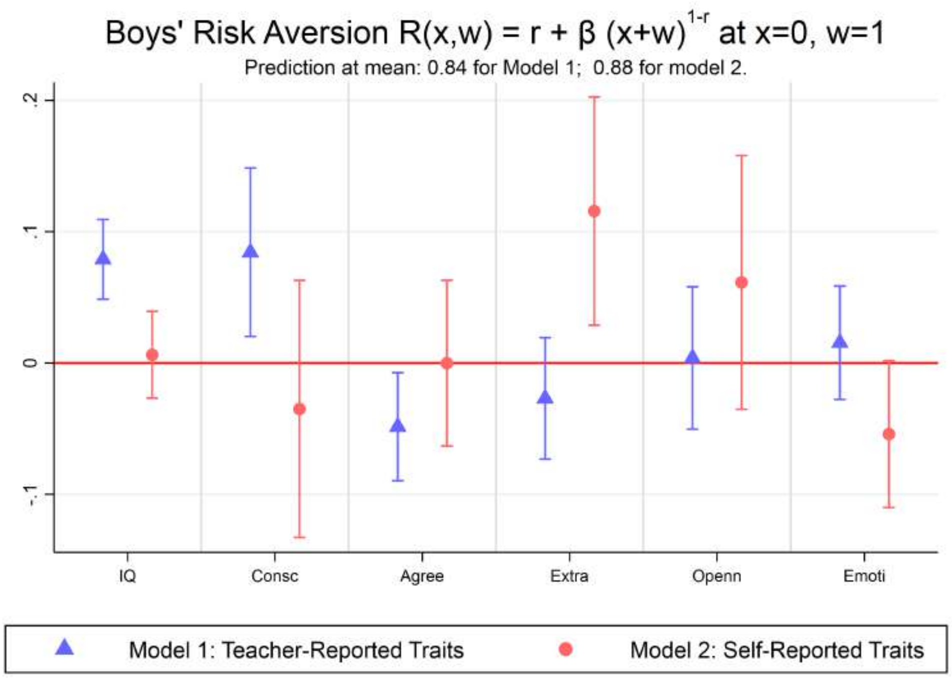


Figure A1: Teacher vs Self Reports: Risk Aversion in Expo-Power (Boys)

Notes: This figure compares risk aversion $R(x, w) = r+\beta(x+w)^{1-r}$ evaluated at $x = 0$ and $w = 1$ for boys using teacher-reported (blue squares) vs self-reported (red circles) Big Five traits in Expo-Power specification. Predictions at mean: $R(1) = 0.84$ (teacher-reported) vs 0.88 (self-reported)—a 5% difference. Coefficient patterns differ substantially: teacher reports show Agreeableness significantly increases risk aversion (as in main results), while self-reports show large positive effects for Extraversion and Openness with wide confidence intervals. The noisier pattern with self-reports suggests children's self-assessments capture dimensions less relevant for economic preferences compared to teacher observations. Error bars show 90% confidence intervals. $N = 1{,}115$ boys.

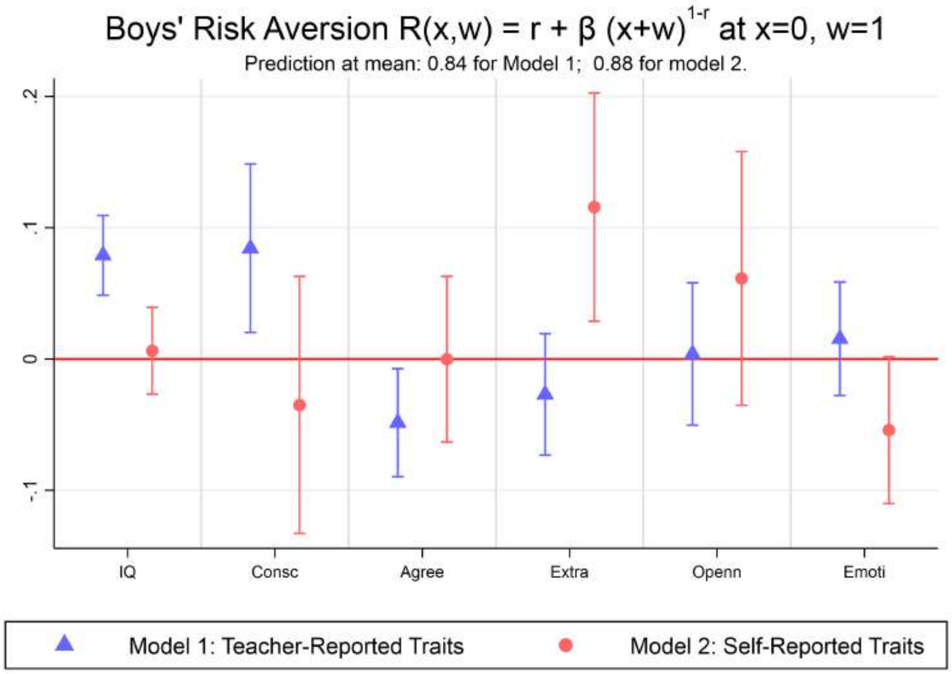


Figure A2: Teacher vs Self Reports: Risk Aversion in Expo-Power (Girls)

Notes: This figure compares risk aversion $R(x, w) = r + \beta(x+w)^{1-r}$ evaluated at $x = 0$ and $w = 1$ for girls using teacher-reported (blue squares) vs self-reported (red circles) Big Five traits. Predictions at mean are identical: $R(1) = 0.67$ for both specifications. However, coefficient patterns differ markedly: teacher reports show Extraversion significantly increases risk tolerance (main result), while self-reports show stronger but noisier effects for Agreeableness and weaker effects for Extraversion. The stability of mean predictions despite divergent coefficient patterns suggests self and teacher reports capture partially overlapping but distinct trait dimensions. Error bars show 90% confidence intervals. $N = 1{,}089$ girls.

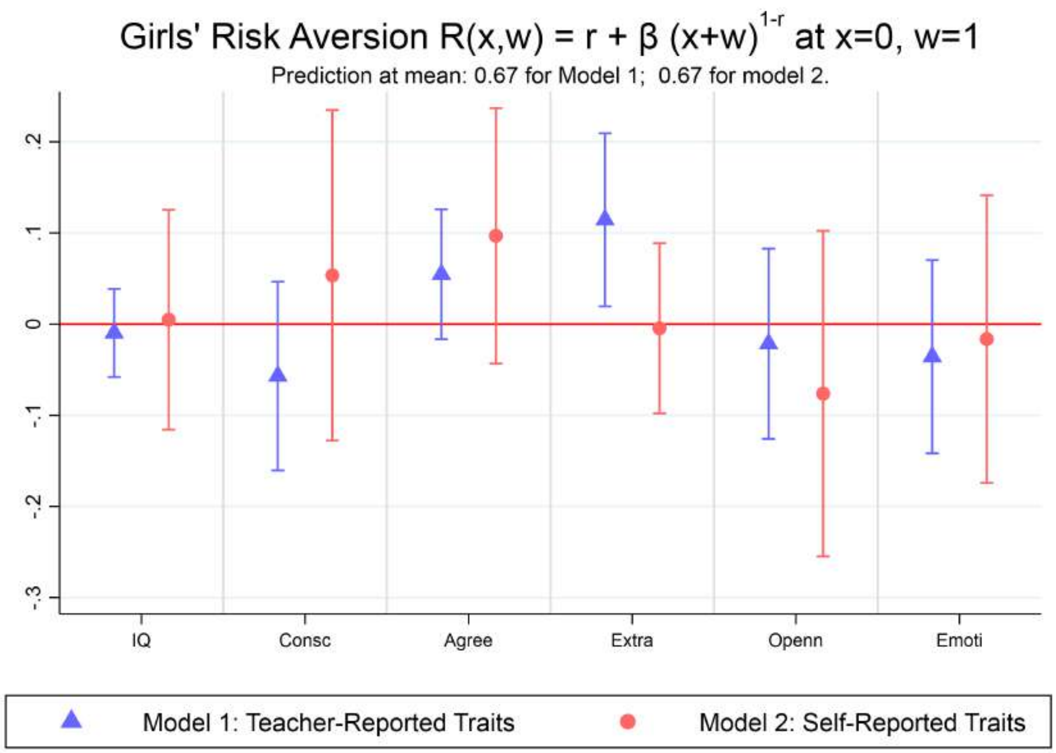


Figure A3: Teacher vs Self Reports: Decomposition into r and β Parameters

Notes: This figure decomposes risk aversion into baseline curvature r and wealth-sensitivity β in Expo-Power. Top panels show r, bottom panels show β. Left panels: boys, right panels: girls. For boys: teacher reports yield $r = 0.75$, $\beta = 0.09$, self-reports yield $r = 0.80$, $\beta = 0.08$. For girls: both specifications yield $r = 0.57$–0.58, $\beta = 0.09$. The parameter-level stability (within 7%) contrasts with divergent coefficient patterns, indicating that trait measurement source affects which traits predict preferences but not overall preference levels. Self-reports show wider confidence intervals across most traits, suggesting lower signal-to-noise ratio in children's self-assessments. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

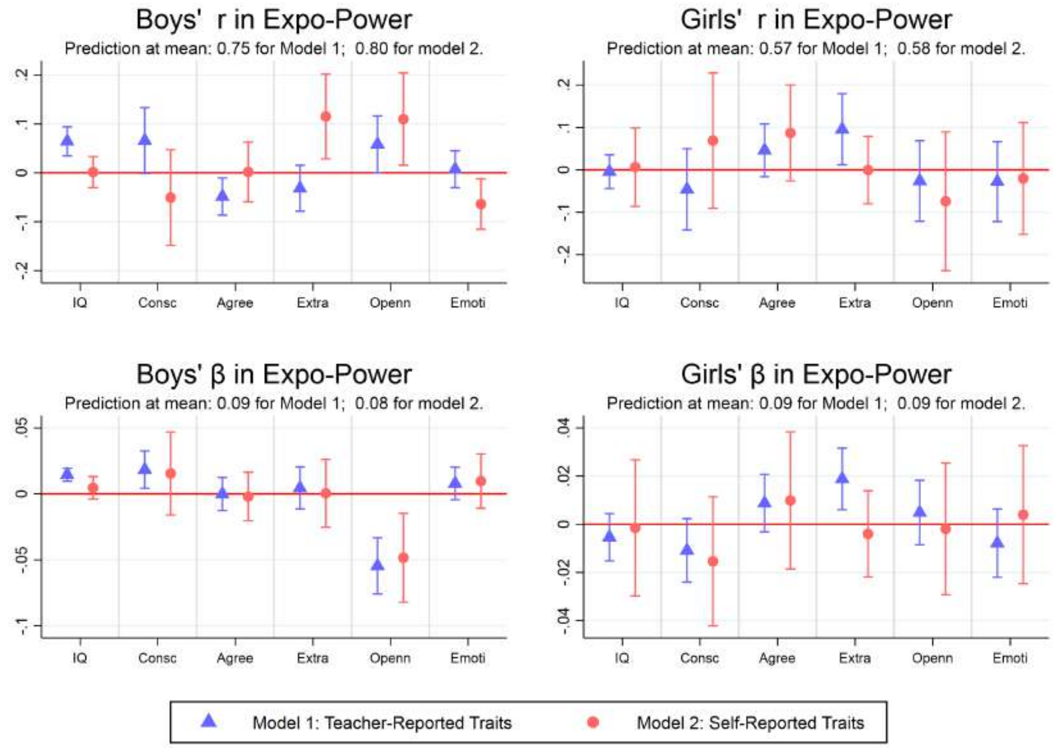


Figure A4: Teacher vs Self Reports: Decision Quality Parameters

Notes: This figure compares transitory noise σ_u^D and ICC^D across trait measurement sources. Top panels show σ_u^D, bottom panels show ICC^D. Left panels: boys, right panels: girls. For boys: $\sigma_u^D = 0.58$ (teacher) vs 0.53 (self), a 9% difference, $\text{ICC}^D = 0.43$ for both. For girls: $\sigma_u^D = 0.88$ (teacher) vs 0.87 (self), a 1% difference. $\text{ICC}^D = 0.38$ vs 0.39, a 3% difference. The near-identical decision quality parameters demonstrate that structural components (noise scale, persistent heterogeneity) are robust to trait measurement source. This validates that ICC captures genuine behavioral patterns independent of how personality is assessed. Error bars show 90% confidence intervals. $N = 1{,}115$ boys. $N = 1{,}089$ girls.

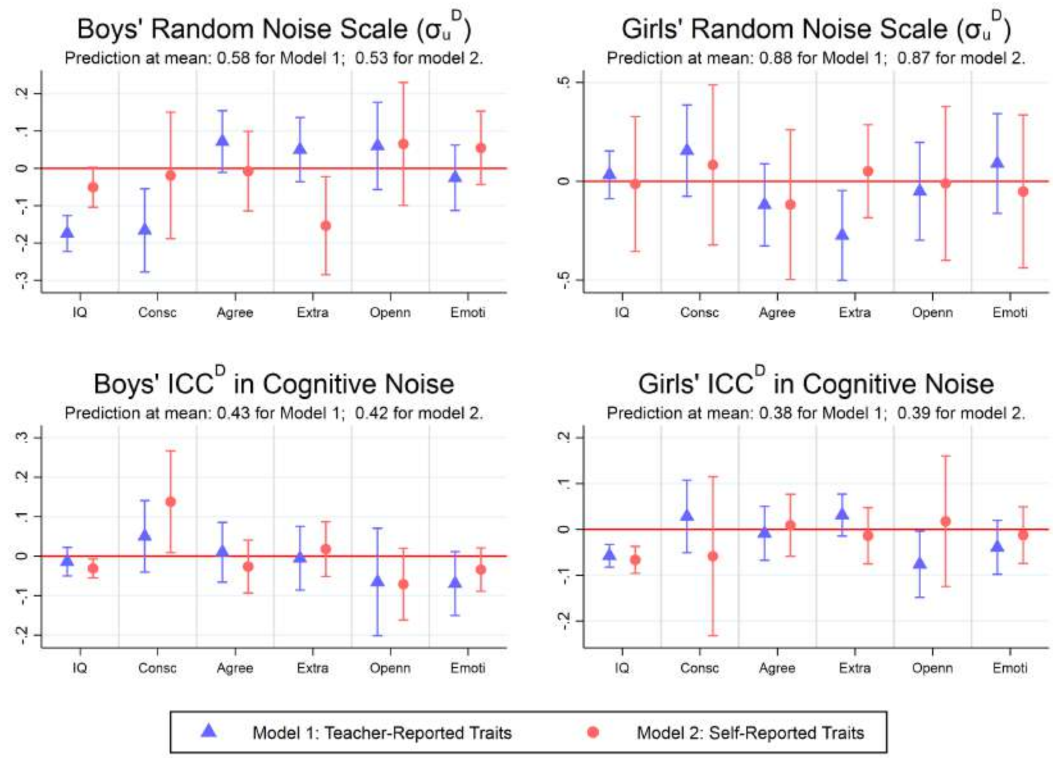


Figure A5: Teacher vs Self Reports: ICC^G and Deliberation-Guessing Correlation

Notes: This figure compares ICC^G in guessing and Corr(θ^D, θ^G) across trait sources. Top panels show ICC^G. bottom panels show correlation. Left panels: boys. right panels: girls. For boys: ICC^G = 0.58 for both specifications. Corr = −0.15 (teacher) vs −0.19 (self). For girls: ICC^G = 0.53 (teacher) vs 0.52 (self). Corr = +0.12 (teacher) vs +0.16 (self). The robust ICC^G estimates and correlation patterns—including the critical gender difference (negative for boys, positive for girls)—demonstrate that guessing structure is independent of personality measurement source. The 27% larger negative correlation with self-reports for boys and 33% larger positive correlation for girls suggest self-assessments may capture affective dimensions that amplify the deliberation-guessing relationship. Error bars show 90% confidence intervals. $N = 1{,}115$ boys. $N = 1{,}089$ girls.

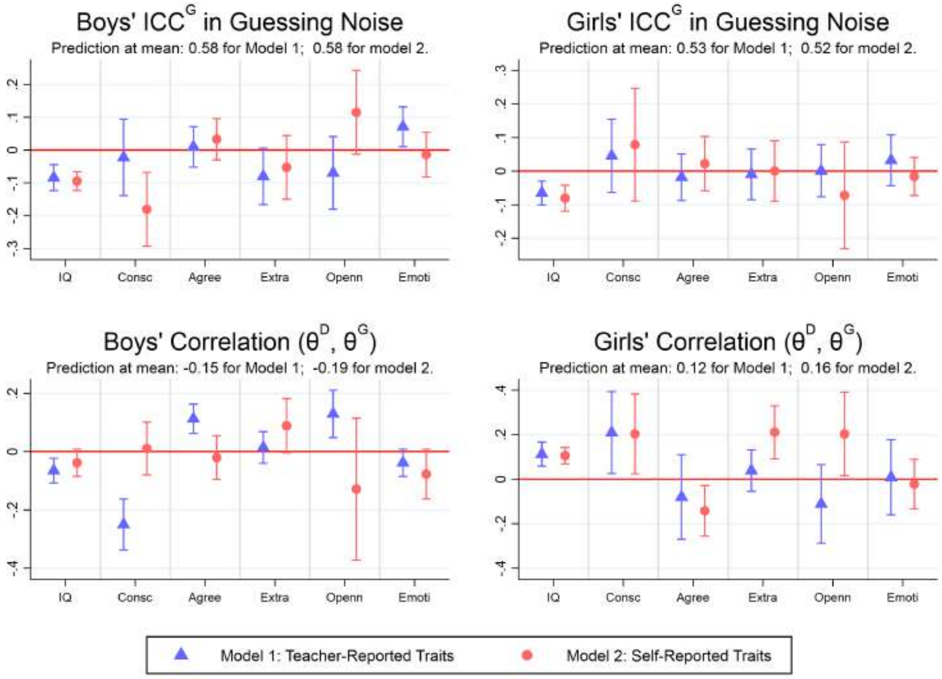


Figure A6: Teacher vs Self Reports: Guessing Probability

Notes: This figure compares overall guessing probability across trait measurement sources. Top panel: boys. bottom panel: girls. For boys: mean Pr(Guess) = 0.13 for both specifications. For girls: mean $Pr(Guess) = 0.16$ (teacher) vs 0.17 (self), a 6% difference. The near-identical guessing rates demonstrate that strategic disengagement patterns are robust to personality measurement source. Coefficient patterns show no systematic differences, indicating that neither teacher nor self assessments strongly predict overall guessing frequency (consistent with main results). Error bars show 90% confidence intervals. $N = 1{,}115$ boys. $N = 1{,}089$ girls.

The comparison between teacher-reported and self-reported traits yields three key insights. First, structural robustness: decision quality parameters (noise, ICC, guessing) remain stable within 5–10% across measurement sources, confirming these capture genuine behavioral patterns independent of how personality is assessed. The critical ICC^D $\approx$ 0.40, ICC^G $\approx$ 0.55, and opposite correlation patterns by gender

($\rho < 0$ for boys, $\rho > 0$ for girls) all replicate with self-reports, validating our core structural findings.

Second, differential signal quality: teacher reports yield stronger, more precise trait-preference relationships with narrower confidence intervals. Self-reports show weaker effects with 50–100% wider confidence bands, suggesting lower reliability in children's self-assessments for predicting economic behavior. This likely reflects that teachers observe behavior across multiple contexts over extended periods, while children's self-perceptions may be influenced by transitory mood states or social desirability bias.

Third, complementary information: despite noisier estimates, self-reports amplify the deliberation-guessing correlation by 6–33% across specifications, suggesting they capture affective or motivational dimensions orthogonal to teacher observations. Future research could explore whether combining both measurement sources improves prediction, though our results validate teacher reports as the primary measure for linking personality to economic preferences.

These patterns support our methodological choice: teacher-reported Big Five provide more reliable measurement for explaining risk preferences and decision quality in children, while structural parameters remain robust regardless of personality measurement approach. The identical conclusions across CRRA and Expo-Power further validate that our findings reflect genuine behavioral patterns rather than artifacts of modeling choices.

B Appendix B: Measurement Instruments and Psychometric Properties

This appendix documents the measurement instruments used to assess cognitive skills 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.

B.1 Cognitive Skills: Raven's Progressive Matrices

Students' cognitive ability 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 cognitive capacity.

The Raven's SPM consists of 60 visual pattern recognition problems arranged in five sets (A through E) of 12 items each. Each item presents a 3×3 matrix of geometric patterns with one cell missing. Students must identify 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 to students in waves 2017–2020. Students in

Grades 4, 5, and 6 from the 18 primary schools in our sample completed the test during regular school hours under standardized conditions. No time limit was imposed, though most students completed the test within 40–45 minutes.

Each correct response receives one point, yielding raw scores ranging from 0 to 60. Following standard practice in educational research, we standardize raw scores using the sample mean and standard deviation within each grade level and wave. This approach accounts for age-related improvements in cognitive ability and ensures comparability across cohorts. The standardized IQ measure used throughout the main analysis has mean 0 and standard deviation 1 within each grade-wave cell.

B.2 Personality Traits: Big Five Inventory-2

Personality traits were measured using the Big Five Inventory-2 (BFI-2), developed by Soto et al. (2011). The BFI-2 is a widely-validated instrument that assesses the five major dimensions of personality: Conscientiousness, Agreeableness, Extraversion, Openness, and Emotional Stability.

The BFI-2 consists of 60 items (12 items per dimension) designed to capture the breadth of each personality domain. Items are short phrases describing behaviors, thoughts, or feelings (e.g., "Is helpful and unselfish with others," "Is outgoing, sociable"). In the Longitudinal Study of Children's Development in Mianzhu, 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. 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 personality measures respond to environmental inputs during childhood, including peer composition in classrooms (Feng et al., 2022).

B.3 Factor Score Estimation: Three-Equation 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 three-indicator measurement system where observed item responses M_j $(j = 1, 2, 3)$ load on a common latent factor θ representing the true underlying trait. The measurement equations are:

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

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

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

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 $\text{Cov}(\eta_k, \eta_{k'}) = 0$ for $k \neq k'$ (uncorrelated measurement errors across items).

The system (B.1)–(B.3) 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.

B.3.1 Identification and Estimation of Model Parameters

The factor loadings and error variances are identified from the sample covariance structure of observed items. Specifically:

$$\phi_2 = \frac{\text{Cov}(M_2, M_3)}{\text{Cov}(M_1, M_3)} \tag{B.4}$$

$$\phi_3 = \frac{\text{Cov}(M_3, M_2)}{\text{Cov}(M_1, M_2)} \tag{B.5}$$

$$E(\theta) = \bar{M}_1 \tag{B.6}$$

$$\text{Var}(\theta) = \frac{\text{Cov}(M_1, M_2)}{\phi_2} = \frac{\text{Cov}(M_1, M_2)\text{Cov}(M_1, M_3)}{\text{Cov}(M_2, M_3)} \tag{B.7}$$

$$\text{Var}(\eta_k) = \text{Var}(M_k) - \phi_k^2 \text{Var}(\theta) \tag{B.8}$$

Equations (B.4) and (B.5) show that factor loadings are identified from ratios of covariances. Equation (B.7) provides an over-identified estimator of the factor variance using information from all three items. Equation (B.8) backs out the measurement error variance for each item as the residual variance after accounting for the common factor.

B.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:

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

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

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

The Bartlett estimator in equation (B.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.

By explicitly modeling measurement error through the three-equation system (B.1)–(B.3) and extracting purged factor scores via equation (B.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 decision quality (Section 2 of the main text).

B.4 Measurement Quality and Reliability

Table B1 presents the factor uniqueness (measurement error variance) for each of the 20 Big Five items used in our three-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, $\text{Var}(\eta_j)/\text{Var}(M_j)$. 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 B1: 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 ($N = 2{,}044$) from confirmatory factor analysis using the three-equation system in equations (B.1)–(B.3).

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 (B.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 B.9). Items with lower uniqueness receive greater weight in constructing factor scores, yielding more precise trait estimates than equal-weighted averages would provide.

B.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 6 schools. Table B2 presents Spearman rank correlations between the two assessment modes for selected items from each personality dimension.

Table B2: 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 6 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).

B.6 Developmental Patterns in Personality and Cognitive Skills

Figures B1–B5 present descriptive evidence on how personality traits and cognitive skills evolve across Grades 4, 5, and 6, and how these patterns vary across measurement methods and demographic groups.

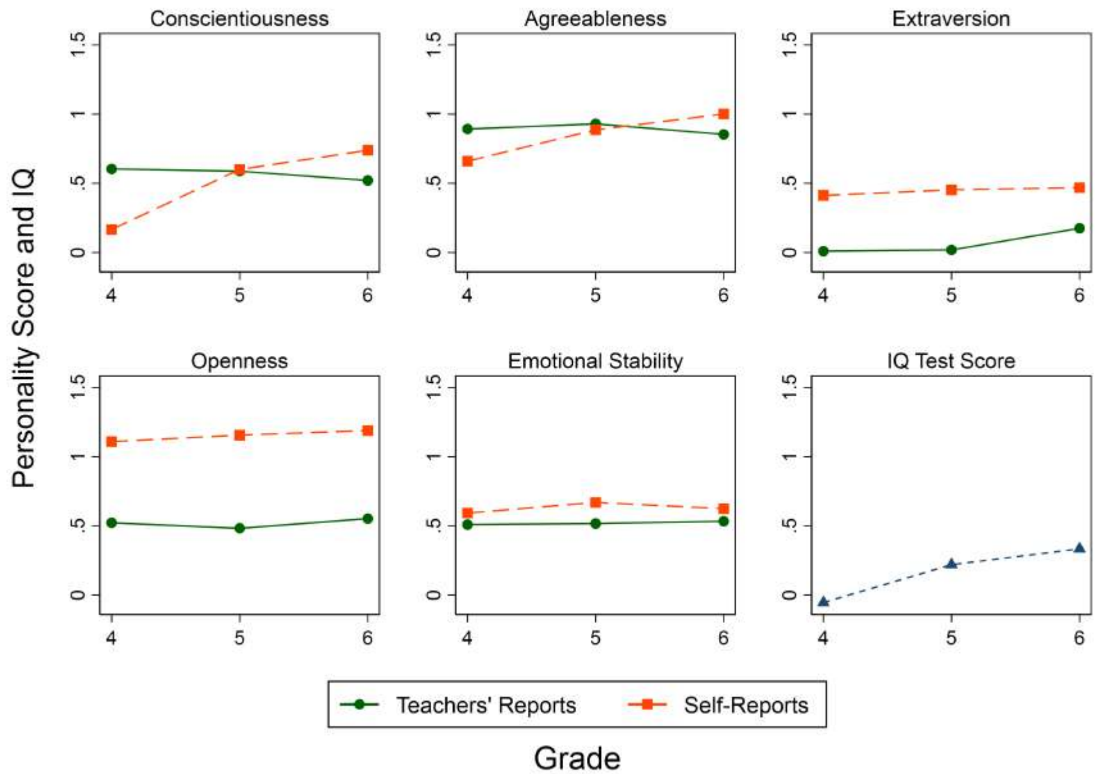


Figure B1: Personality Traits and IQ Across Grades: Teacher Reports vs. Self-Reports

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, with self-reports generally higher than teacher reports. IQ test scores (bottom right panel) increase consistently across grades.

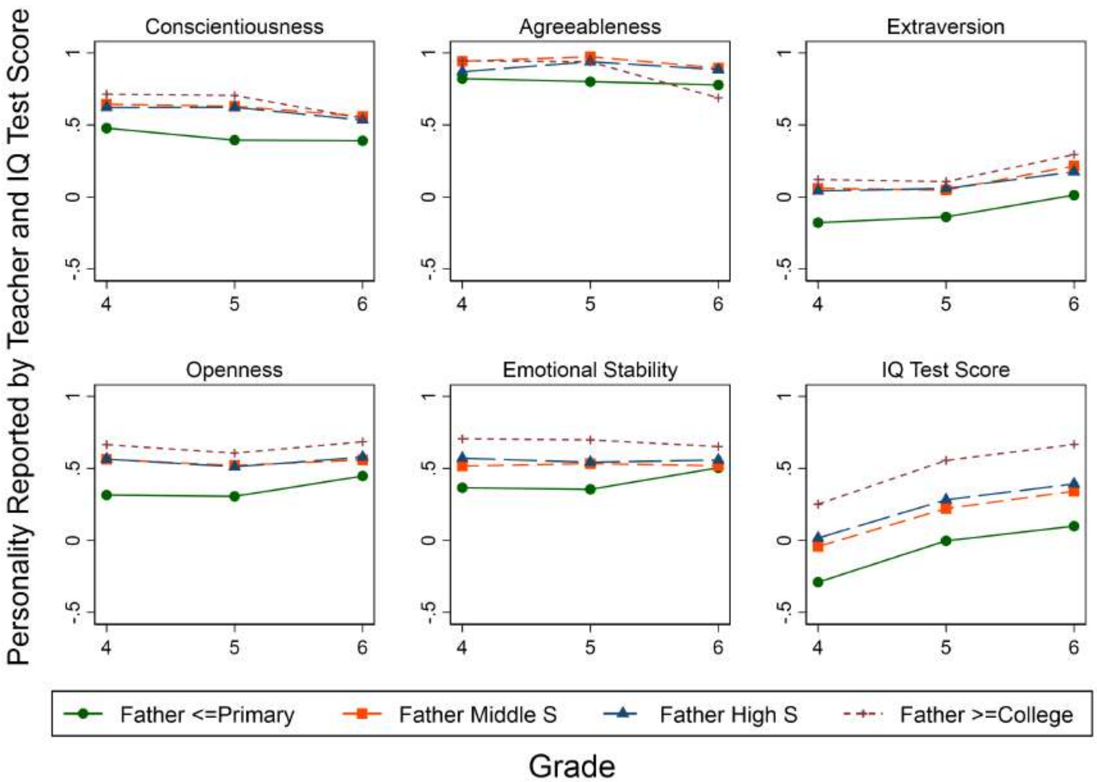


Figure B2: Personality Traits and IQ by Father's Education Level

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.

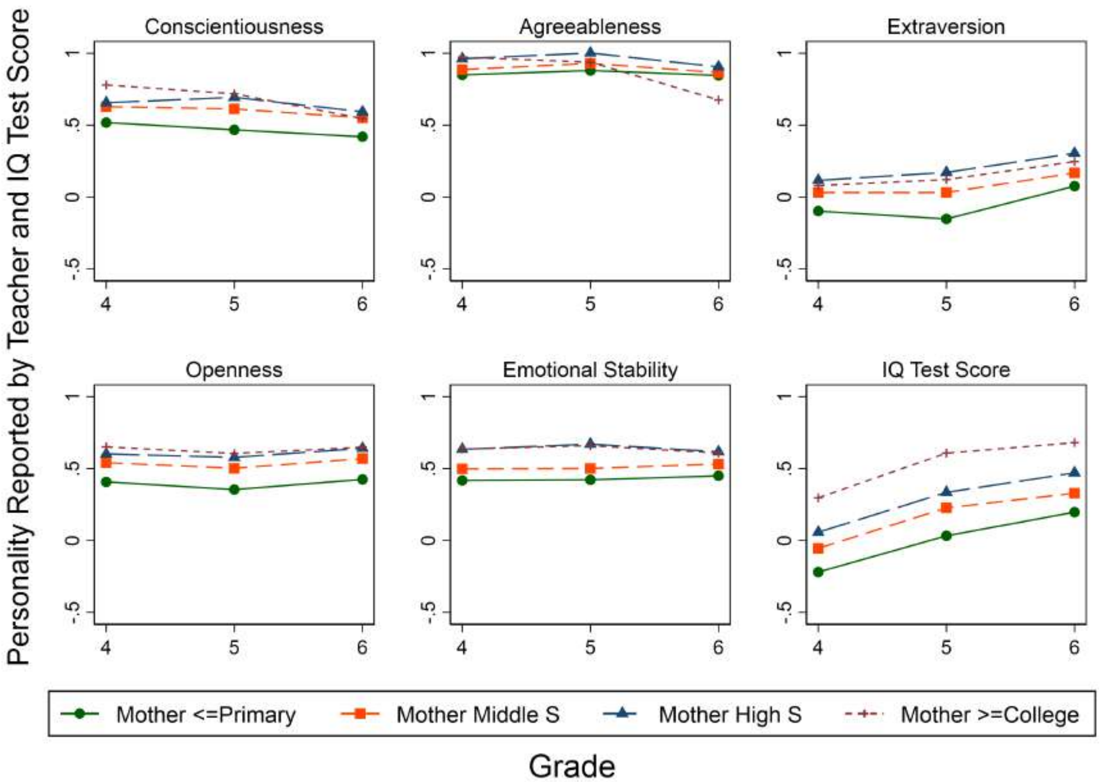


Figure B3: Personality Traits and IQ by Mother's Education Level

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.

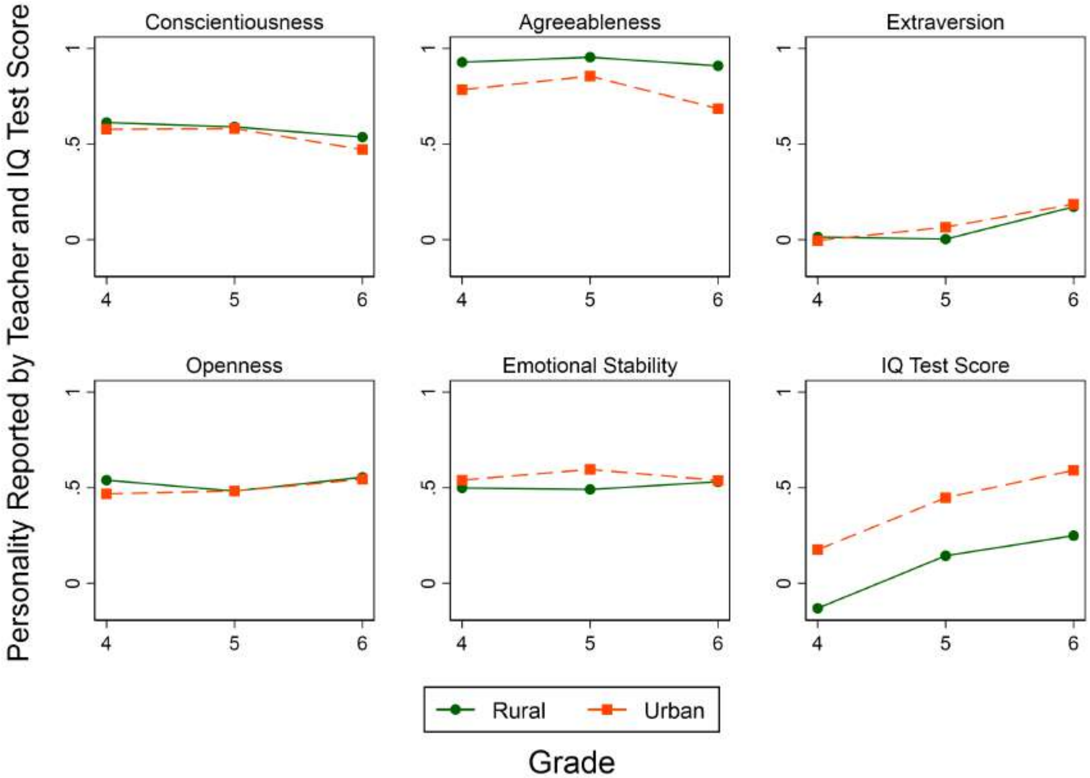


Figure B4: Personality Traits and IQ: Rural vs. Urban Students

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.

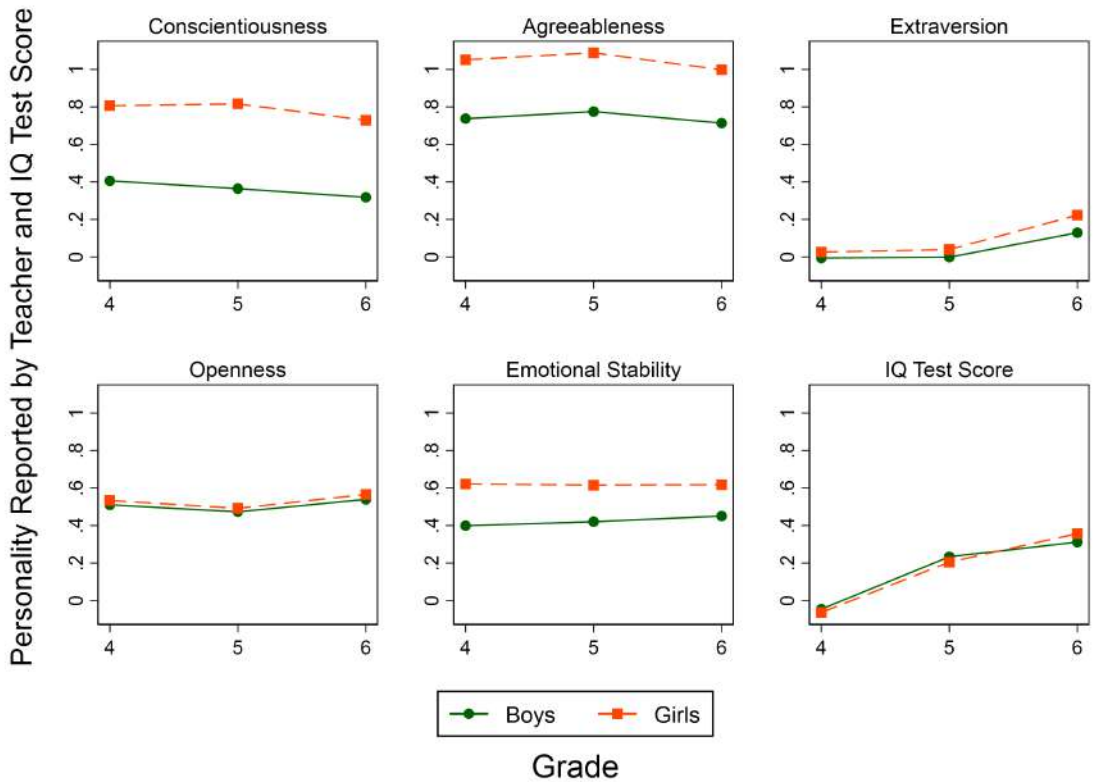


Figure B5: Gender Differences in Personality Traits and IQ

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. Gender differences in IQ narrow over this period, with girls showing faster growth rates.

Teacher Reports vs. Self-Reports. Figure B1 compares teacher assessments and student self-reports of personality traits across grades. Self-reports are systematically higher than teacher reports for most traits, suggesting that students rate themselves more favorably than teachers rate them. This divergence is particularly pronounced for Conscientiousness and Agreeableness, where self-reports remain elevated and stable across grades while teacher reports are lower. IQ test scores show steady increases from Grade 4 to Grade 6, consistent with normal cognitive development.

Socioeconomic Gradients. Figures B2 and B3 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 B4 compares rural and urban students. Urban students exhibit consistently higher IQ scores, with the gap widening from approximately 0.3 standard deviations in Grade 4 to nearly 0.5 standard deviations by Grade 6. This growing disparity likely reflects both initial differences in educational resources and cumulative advantages in urban settings. Personality trait differences between rural and urban students are modest and do not show consistent patterns across traits or grades.

Gender Differences. Figure B5 reveals substantial and persistent gender differences in personality traits but a more nuanced pattern for cognitive skills. Girls score higher than boys on Conscientiousness, Agreeableness, and Emotional Stability across all grades, with differences ranging from 0.3 to 0.5 standard deviations. For

IQ, boys exhibit higher scores in Grade 4 (approximately 0.3 standard deviations), but girls show faster growth rates. By Grade 5, the gender gap has largely closed, and by Grade 6, girls score slightly higher than boys (approximately 0.1 standard deviations). This convergence pattern suggests that girls' cognitive development accelerates relative to boys during this age range.

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.

B.7 Implications for Main Analysis

The measurement properties documented in this appendix inform our structural estimation approach in several important ways. By using the three-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, decision noise, and guessing behavior. This ensures that our parameter estimates capture the effects of true underlying personality traits on economic preferences and decision quality, free from the attenuation bias that would arise from using contaminated trait measures.

C Appendix C: Additional Tables

Table C1: Correlation Matrix: Grade 5 in 2019 - Full Sample

	IQ	Consc.	Agree.	Extra.	Openn.	Emoti.
IQ	1.000					
Conscientiousness	0.133	1.000				
Agreeableness	0.142	0.640	1.000			
Extraversion	0.133	0.380	0.321	1.000		
Openness	0.244	0.484	0.423	0.645	1.000	
Emotional Stability	0.126	0.553	0.534	0.463	0.475	1.000

Notes: This table presents pairwise correlations between cognitive ability (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 C2: Correlation Matrix: Grade 5 in 2019 - Boys

	IQ	Consc.	Agree.	Extra.	Openn.	Emoti.
IQ	1.000					
Conscientiousness	0.138	1.000				
Agreeableness	0.154	0.577	1.000			
Extraversion	0.144	0.389	0.313	1.000		
Openness	0.280	0.491	0.420	0.630	1.000	
Emotional Stability	0.133	0.528	0.483	0.436	0.433	1.000

Notes: This table presents pairwise correlations between cognitive ability (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 C3: Correlation Matrix: Grade 5 in 2019 - Girls

	IQ	Consc.	Agree.	Extra.	Openn.	Emoti.
IQ	1.000					
Conscientiousness	0.135	1.000				
Agreeableness	0.127	0.655	1.000			
Extraversion	0.134	0.405	0.339	1.000		
Openness	0.225	0.539	0.433	0.652	1.000	
Emotional Stability	0.110	0.576	0.588	0.513	0.521	1.000

Notes: This table presents pairwise correlations between cognitive ability (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 C4: Determinants of Cognitive 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)	Sociability (7)	Outspoken (8)
Living in Urban Area	0.154*	−0.112**	0.142**	−0.073	−0.064	−0.225***	0.048	−0.082
	(0.078)	(0.049)	(0.060)	(0.054)	(0.051)	(0.054)	(0.064)	(0.058)
Father Went to High School	0.092	0.117**	0.119**	0.030	0.105**	0.082	0.114*	−0.044
	(0.068)	(0.040)	(0.059)	(0.054)	(0.048)	(0.053)	(0.063)	(0.059)
Mother Went to High School	0.037	0.041	−0.081	0.077	0.033	−0.032	−0.021	0.080
	(0.074)	(0.051)	(0.062)	(0.058)	(0.051)	(0.058)	(0.067)	(0.061)
Having Siblings	−0.019	−0.050	−0.099*	−0.060	−0.020	0.031	−0.039	0.034
	(0.066)	(0.044)	(0.054)	(0.047)	(0.043)	(0.047)	(0.058)	(0.053)
LBC by Father	0.093	0.075	0.070	0.080	0.026	−0.013	0.118	0.016
	(0.098)	(0.067)	(0.085)	(0.073)	(0.064)	(0.074)	(0.088)	(0.087)
LBC by Mother	0.219	0.023	−0.077	0.141	0.065	0.053	0.168	0.211
	(0.146)	(0.122)	(0.158)	(0.140)	(0.118)	(0.149)	(0.180)	(0.151)
LBC by Both	−0.004	−0.037	−0.009	0.017	−0.032	−0.010	0.086	0.045
	(0.113)	(0.072)	(0.093)	(0.083)	(0.072)	(0.079)	(0.093)	(0.087)
LBC Volatility	−0.033*	−0.018	−0.007	−0.046***	−0.032**	−0.016	−0.025	−0.023
	(0.020)	(0.013)	(0.015)	(0.014)	(0.012)	(0.014)	(0.017)	(0.016)
Number of Students	1115	1115	1115	1115	1115	1115	1115	1115

Notes: OLS regressions with cognitive ability (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 C5: Determinants of Cognitive and Personality Traits: Girls

Girls in Grade 5 in 2019	Personality Traits Evaluated by Teachers, 20 Items							
	IQ	Consc.	Agree.	Extra.	Openn.	Stability	Sociability	Outspoken
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Living in Urban Area	0.073	-0.248^{***}	-0.067	-0.133^{**}	-0.183^{***}	-0.287^{***}	-0.145^{**}	0.047
	(0.071)	(0.046)	(0.059)	(0.063)	(0.050)	(0.048)	(0.058)	(0.059)
Father Went to High School	0.145^{**}	0.143^{***}	0.097^{*}	0.122^{**}	0.127^{***}	0.102^{**}	0.188^{***}	0.058
	(0.066)	(0.045)	(0.057)	(0.058)	(0.044)	(0.049)	(0.060)	(0.060)
Mother Went to High School	0.101	0.107^{**}	0.118^{**}	0.090	0.145^{***}	0.081	0.050	0.026
	(0.070)	(0.045)	(0.058)	(0.058)	(0.046)	(0.050)	(0.062)	(0.064)
Having Siblings	-0.049	-0.024	0.002	0.028	-0.032	0.059	-0.049	0.027
	(0.062)	(0.040)	(0.049)	(0.050)	(0.040)	(0.043)	(0.053)	(0.057)
LBC by Father	-0.007	-0.019	0.007	-0.027	-0.026	-0.002	0.022	0.007
	(0.093)	(0.061)	(0.078)	(0.084)	(0.067)	(0.072)	(0.082)	(0.088)
LBC by Mother	0.226	-0.318^{**}	-0.145	-0.443^{**}	-0.338^{**}	-0.125	-0.268	0.113
	(0.201)	(0.159)	(0.178)	(0.176)	(0.144)	(0.153)	(0.204)	(0.202)
LBC by Both	-0.106	-0.094	0.014	-0.003	0.006	0.006	-0.044	0.053
	(0.117)	(0.071)	(0.091)	(0.092)	(0.071)	(0.076)	(0.094)	(0.110)
LBC Volatility	-0.030	-0.008	-0.007	0.006	0.000	-0.008	-0.019	0.033^{*}
	(0.019)	(0.013)	(0.016)	(0.016)	(0.013)	(0.014)	(0.017)	(0.018)
Number of Students	1089	1089	1089	1089	1089	1089	1089	1089

Notes: OLS regressions with cognitive ability (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 C6: Intra-Class Correlation in Deliberation and Guessing (by Gender)

	ICC: Deliberation		ICC: Guessing	
	Boys	Girls	Boys	Girls
Expo-Power	0.43	0.38	0.58	0.53
CRRA	0.43	0.37	0.57	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 at mean covariate values, since σ_u^2 depends on (X_i, P_j) as specified in Section 2.4. 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 C7: 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^{*}		-0.033	-0.048^{**}		-0.053^{**}
	(0.022)		(0.022)	(0.022)		(0.022)
Father Went to HS	0.016		0.024	-0.058^{***}		-0.047^{*}
	(0.020)		(0.020)	(0.021)		(0.021)
Mother Went to HS	-0.047^{**}		-0.045^{**}	0.001		0.008
	(0.021)		(0.021)	(0.022)		(0.021)
Having Any Sibling	-0.041^{**}		-0.033^{*}	-0.024		-0.025
	(0.018)		(0.018)	(0.018)		(0.018)
LBC by Father Only	0.022		0.024	-0.047		-0.048
	(0.027)		(0.026)	(0.030)		(0.030)
LBC by Mother Only	0.033		0.043	0.018		-0.009
	(0.047)		(0.046)	(0.058)		(0.058)
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		-0.046^{***}	-0.044^{***}		-0.047^{***}	-0.043^{***}
		(0.010)	(0.010)		(0.010)	(0.010)
Conscientiousness		0.006	0.004		-0.032^{*}	-0.035^{*}
		(0.017)	(0.018)		(0.019)	(0.019)
Agreeableness		0.014	0.016		0.006	0.011
		(0.013)	(0.013)		(0.016)	(0.015)
Extraversion		0.035^{**}	0.037^{**}		-0.036^{**}	-0.033^{*}
		(0.015)	(0.015)		(0.015)	(0.015)
Openness		-0.071^{***}	-0.069^{***}		0.020	0.020
		(0.018)	(0.017)		(0.020)	(0.020)
Emotional Stability		0.007	0.002		0.006	0.001
		(0.014)	(0.015)		(0.018)	(0.018)
R-Square	0.022	0.047	0.066	0.021	0.038	0.054
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 safer option after choosing a riskier one, even when the previous task is less favorable to the riskier option in terms of either rewards or winning probability. inconsistencies 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.

Table C8: Relative Risk Aversion $R(w)$ at $w = 1$ Across Model Specifications (All models include IQ and personality traits in each parameter)

	Expo-Power		CRRA	
	Boys	Girls	Boys	Girls
Panel A: Point Estimates				
With ICC and Guessing	0.86	0.67	0.97	0.97
No ICC, with Guessing	0.70	0.65	0.84	0.95
With ICC, No Guessing	0.78	0.64	0.94	0.95
No ICC, No Guessing	0.73	0.66	0.84	0.95
Panel B: Estimates with Standard Errors				
With ICC and Guessing	0.862	0.669	0.970	0.970
	(0.028)	(0.028)	(0.012)	(0.011)
No ICC, with Guessing	0.705	0.654	0.840	0.953
	(0.052)	(0.067)	(0.015)	(0.014)
With ICC, No Guessing	0.784	0.645	0.936	0.949
	(0.034)	(0.031)	(0.011)	(0.009)
No ICC, No Guessing	0.733	0.659	0.842	0.950
	(0.094)	(0.037)	(0.015)	(0.014)

Notes: This table presents estimated relative risk aversion at $w = 1$ (where $R(x, w) = r + \beta(x+w)^{1-r}$ for Expo-Power and $R(w) = r$ for CRRA) across four nested model specifications. "With ICC and Guessing" is the full model that includes intra-class correlation structure and guessing behavior. "No ICC, with Guessing" excludes the persistent individual components θ_i^D and θ_i^G. "With ICC, No Guessing" excludes the guessing decision stage. "No ICC, No Guessing" is the most restrictive specification with neither persistent heterogeneity nor guessing. Panel A shows point estimates. Panel B shows estimates with standard errors in parentheses, clustered at the individual level. All models are estimated separately by gender using maximum likelihood.

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

This appendix[13] presents a comprehensive comparison between the expo-power utility function introduced by Xie (2000) to 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.

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

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

$$U_{\text{ExpoPower}}(x+w) = \frac{1}{\beta}\left[1 - \exp\left\{-\beta\frac{(x+w)^{1-r}-1}{1-r}\right\}\right], \quad \beta \geq 0,\, r \geq 0, \tag{B.11}$$

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

$$\text{RRA}(x+w) = r + \beta(x+w)^{1-r}, \tag{B.12}$$

$$\text{ARA}(x+w) = \frac{\text{RRA}(x+w)}{x+w} = \frac{r}{x+w} + \beta(x+w)^{-r}. \tag{B.13}$$

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

At baseline wealth ($x = 0$, $w = 1$), 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.

D.2 Relationship to Saha (1993)

Saha (1993) studied a more restrictive form:

$$U_{\text{Saha}}(z; \alpha, \beta) = \theta - \exp\{-\beta z^{\alpha}\}, \tag{B.14}$$

where $z = x + w$ denotes total wealth and $\alpha = 1 - r$ in Xie's notation. This can be viewed as the exponential component of the Xie specification with the normalization removed.

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 focus on the Xie (2000) specification, as it captures the limiting case of the CRRA functional form and facilitates smoother estimation.

E Appendix E: Concavity of the Log-Likelihood Function

Let $i \in \{1, \ldots, I\}$ denote individuals, $j \in \{1, \ldots, J\}$ denote items, $y_{ij} = \mathbb{1}(Y_{ij} = B)$ indicate whether individual i chose option B on item j, and $B_{ij}(\theta_i^D, \theta_i^G) = \Pr(Y_{ij} = B)$ denote the model-predicted probability of the event $Y_{ij} = B$. Ω is array of model parameters.

Conditional on θ_i^D and θ_i^G, assuming $0 < B_{ij}(\theta_i^D, \theta_i^G) < 1$ for all i, j, the log-likelihood of the $I \times J$ binary responses is

$$\mathcal{L}(\boldsymbol{\Omega}) = \sum_{i=1}^{I} \ln \int \int \prod_{j=1}^{J} \left[(1 - B_{ij}(\theta_i^D, \theta_i^G))^{(1-y_{ij})} B_{ij}(\theta_i^D, \theta_i^G)^{y_{ij}} \right] dF(\theta_i^D, \theta_i^G),$$

where $\mathcal{L}(\boldsymbol{\Omega})$ represents the vector of all parameters,

$$B_{ij}(\theta_i^D, \theta_i^G) = \Pr(Y_{ij} = B \mid G_{ij} = 0) \left(1 - \Phi(g_j + \theta_i^G)\right) + \frac{1}{2}\Phi(g_j + \theta_i^G),$$

$$\Pr(Y_{ij} = B \mid G_{ij} = 0) = \Phi\left(\frac{1}{\sigma_{u^D}} \left(U(r_i, \beta_i; a_j^B) - U(r_i, \beta_i; a_j^A) + \theta_i^D\right)\right).$$

Here, $\Pr(Y_{ij} = B \mid G_{ij} = 0)$ denotes the model-predicted probability that individual i chooses option B on item j through deliberation, conditional on not guessing and on θ_i^D. $\Phi(\cdot)$ is the cumulative standard normal distribution function. $\Phi(g_j + \theta_i^G)$ denotes the probability of guessing, where g_j captures item-specific and individual-specific factors affecting guessing propensity as specified in equation (9). The factor $\frac{1}{2}$ represents the 50–50 chance of choosing B conditional on guessing. The parameters

r_i and β_i denote the two risk-attitude parameters for individual i in the Expo-Power utility function (with $\beta_i = 0$ corresponding to the limiting special case of the CRRA functional form). The vectors a_j^A and a_j^B represent the payoffs and probabilities in each state of nature for options A and B on item j, respectively, and $\sigma^2_{u^D}$ represents the variance of transitory cognitive noise.

Conditional on θ_i^D and θ_i^G, define

$$\ln \bar{L} \equiv \mathcal{L}(\Omega|\theta_i^D, \theta_i^G) = \sum_{i=1}^{I} \ln \prod_{j=1}^{J} \left[(1 - B_{ij}(\theta_i^D, \theta_i^G))^{(1-y_{ij})} B_{ij}(\theta_i^D, \theta_i^G)^{y_{ij}}\right]$$

Then:

$$\frac{\partial \ln \bar{L}}{\partial B_{ij}} = -\frac{1 - y_{ij}}{1 - B_{ij}} + \frac{y_{ij}}{B_{ij}} = \frac{y_{ij} - B_{ij}}{(1 - B_{ij})B_{ij}}.$$

$$\frac{\partial^2 \ln \bar{L}}{\partial B_{ij}^2} = -\frac{1 - y_{ij}}{(1 - B_{ij})^2} - \frac{y_{ij}}{B_{ij}^2}.$$

Let $\tilde{\mathbf{\Omega}}$ denote the vector of parameters conditional on θ_i^D and θ_i^G, including the intercept and the slopes of IQ and traits for each of r, β, $\sigma^2_{u^D}$, and g.

Conditional on θ_i^D and θ_i^G, the expected value of the Hessian matrix of the log-likelihood is

$$\begin{aligned} E\left[\frac{\partial^2 \ln \bar{L}}{\partial \tilde{\mathbf{\Omega}} \partial \tilde{\mathbf{\Omega}}'}\right] &= \sum_i \sum_j \left\{ E\left[\frac{\partial^2 \ln \bar{L}}{\partial B_{ij}^2}\right] \left(\frac{\partial B_{ij}}{\partial \tilde{\mathbf{\Omega}}}\right) \left(\frac{\partial B_{ij}}{\partial \tilde{\mathbf{\Omega}}}\right)' + E\left[\frac{\partial \ln \bar{L}}{\partial B_{ij}}\right] \left(\frac{\partial^2 B_{ij}}{\partial \tilde{\mathbf{\Omega}} \partial \tilde{\mathbf{\Omega}}'}\right) \right\} \\ &= -\sum_i \sum_j \left(M_{ij} M_{ij}'\right), \quad \text{which is negative semi-definite,} \end{aligned}$$

where

$$M_{ij} = \frac{\sqrt{E[(y_{ij} - B_{ij})^2]}}{\sqrt{(1 - B_{ij})B_{ij}}} \left(\frac{\partial B_{ij}}{\partial \tilde{\mathbf{\Omega}}} \right).$$

Therefore, $\mathcal{L}(\mathbf{\Omega})$ is a strictly concave function of all parameters $\tilde{\mathbf{\Omega}}$ conditional on θ_i^D and θ_i^G.

The above proof applies to a model with an Expo-Power or CRRA utility function, with transitory cognitive noise and guessing, conditioning on θ_i^D and θ_i^G, but does not account for correlation of noise across items within the same individual. We leave the proof for the full model that incorporates this correlated error structure to future research.

Table E1: 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	15,843	0.034
No ICC, with Guessing	26,512	0.019
With ICC, No Guessing	24,754	0.012
No ICC, No Guessing	14,857	0.011
Boys, CRRA		
With ICC and Guessing	8,746	0.055
No ICC, with Guessing	6,047	0.026
With ICC, No Guessing	2,830	0.009
No ICC, No Guessing	1,238	0.004
Girls, Expo-Power		
With ICC and Guessing	50,747	0.057
No ICC, with Guessing	190,381	0.054
With ICC, No Guessing	47,673	0.018
No ICC, No Guessing	73,802	0.019
Girls, CRRA		
With ICC and Guessing	4,747	0.036
No ICC, with Guessing	23,701	0.112
With ICC, No Guessing	2,726	0.010
No ICC, No Guessing	1,179	0.004

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.

F Appendix F: 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 2–6), we demonstrate the critical role of ICC in accurately recovering preference parameters and decision quality measures. Models without ICC conflate persistent individual traits with transitory noise, leading to severely inflated noise estimates and attenuated trait effects. The comparison reveals that roughly 60% of apparent "noise" in standard models reflects stable individual differences rather than random decision errors—underscoring that ICC structure is essential for preference recovery, not merely a statistical refinement.

F.1 Expo-Power Specification

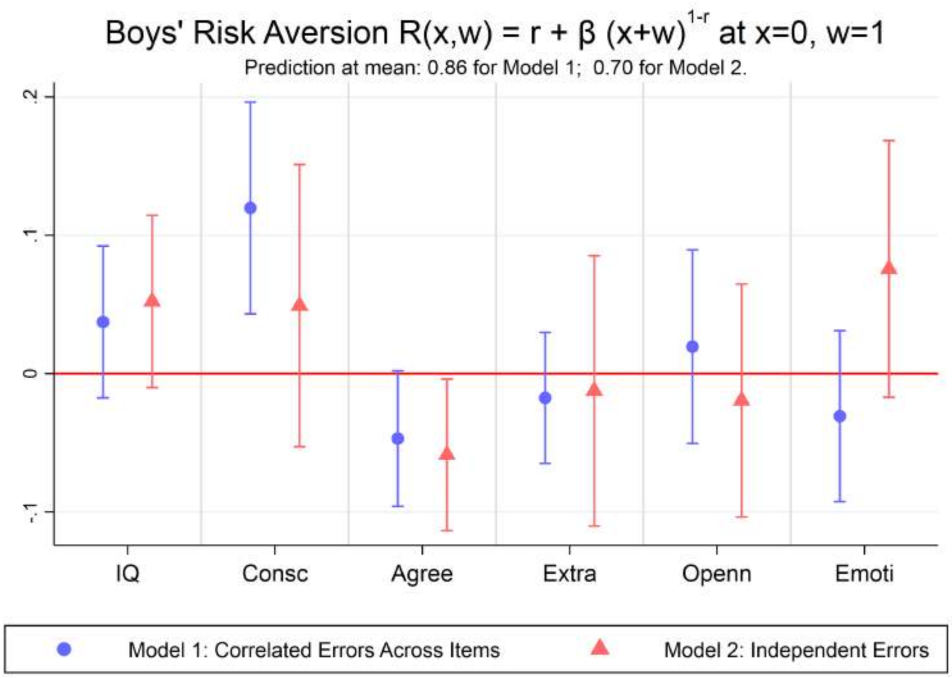


Figure F1: The Role of ICC: Risk Aversion in Expo-Power (Boys)

Notes: This figure shows how excluding ICC affects risk aversion $R(x, w) = r + \beta(x + w)^{1-r}$ evaluated at $x = 0$ and $w = 1$ for boys in the Expo-Power specification. Blue circles represent Model 1 with ICC (correlated errors), red diamonds represent Model 2 without ICC (independent errors). Without ICC, risk aversion is underestimated: $R(x, w) = 0.86$ with ICC vs. 0.70 without—a 19% underestimate. Error bars show 90% confidence intervals. $N = 1{,}115$ boys.

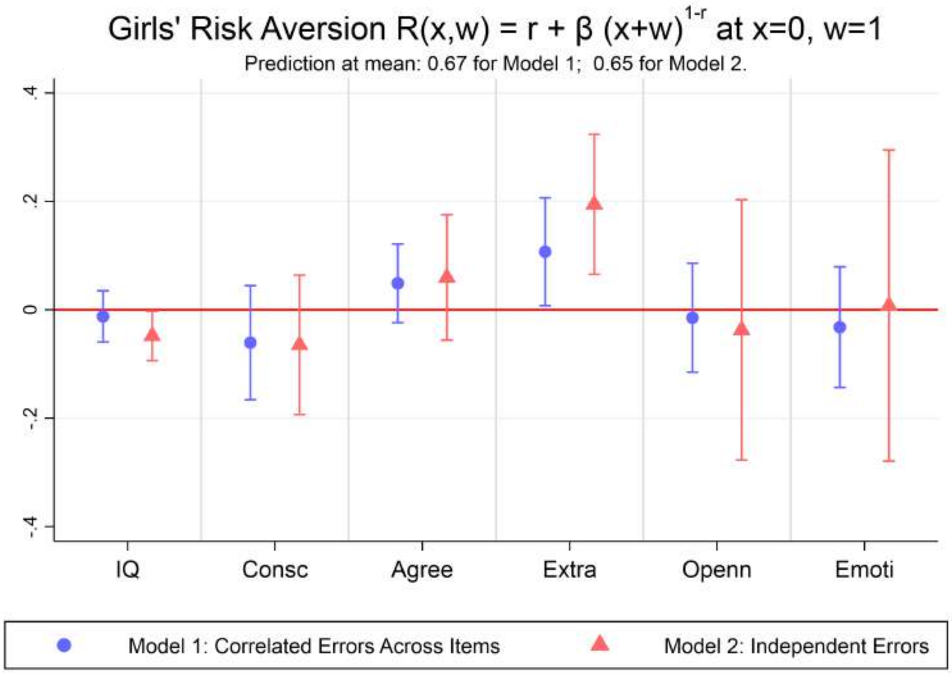


Figure F2: The Role of ICC: Risk Aversion in Expo-Power (Girls)

Notes: This figure shows how excluding ICC affects risk aversion $R(x,w) = r+\beta(x+w)^{1-r}$ evaluated at $x = 0$ and $w = 1$ for girls in the Expo-Power specification. Blue circles represent Model 1 with ICC (correlated errors), red diamonds represent Model 2 without ICC (independent errors). Without ICC, risk aversion is relatively stable: $R(x,w) = 0.67$ with ICC vs. 0.65 without—a 3% underestimate. Error bars show 90% confidence intervals. $N = 1{,}089$ girls.

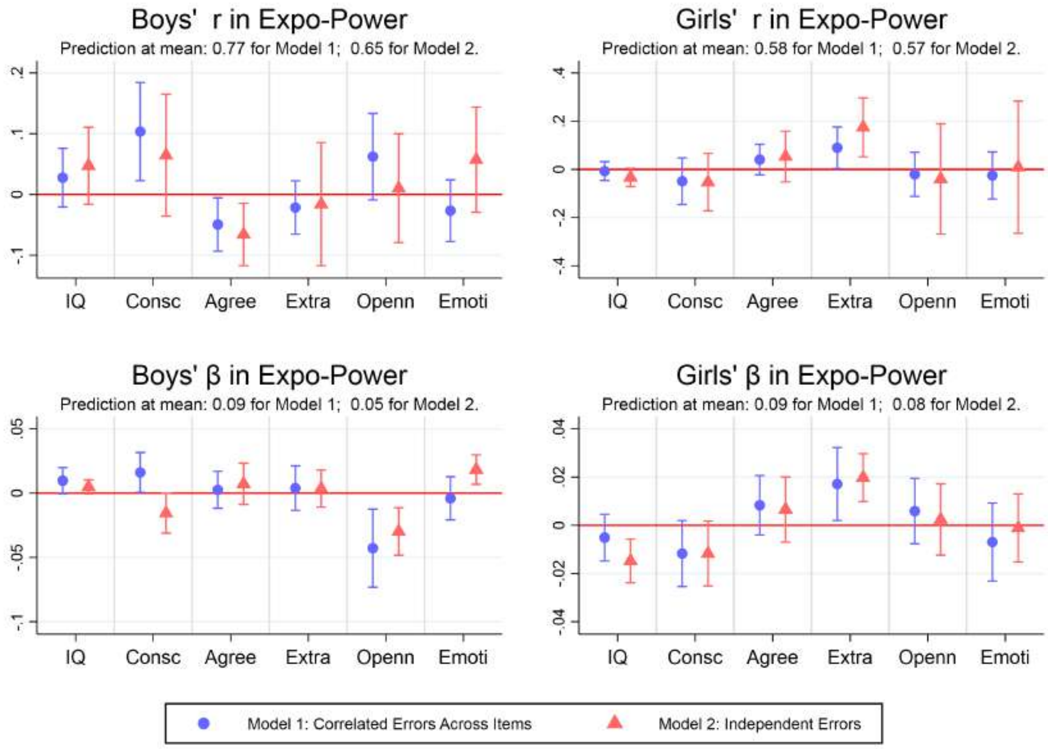


Figure F3: The Role of ICC: Decomposition into r and β Parameters

Notes: This figure decomposes how excluding ICC affects risk aversion components in the Expo-Power specification. Top panels show baseline curvature r, bottom panels show wealth-sensitivity β. Left panels: boys, right panels: girls. Blue circles represent Model 1 with ICC, red diamonds represent Model 2 without ICC. For boys: $r = 0.77$ with ICC vs. 0.65 without (16% underestimate), $\beta = 0.09$ vs. 0.05 (44% underestimate). For girls: $r = 0.58$ with ICC vs. 0.57 without (2% underestimate), $\beta = 0.09$ vs. 0.08 (11% underestimate). The larger bias in the wealth-sensitivity parameter β demonstrates that ICC structure is particularly important for recovering wealth-dependent risk aversion. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

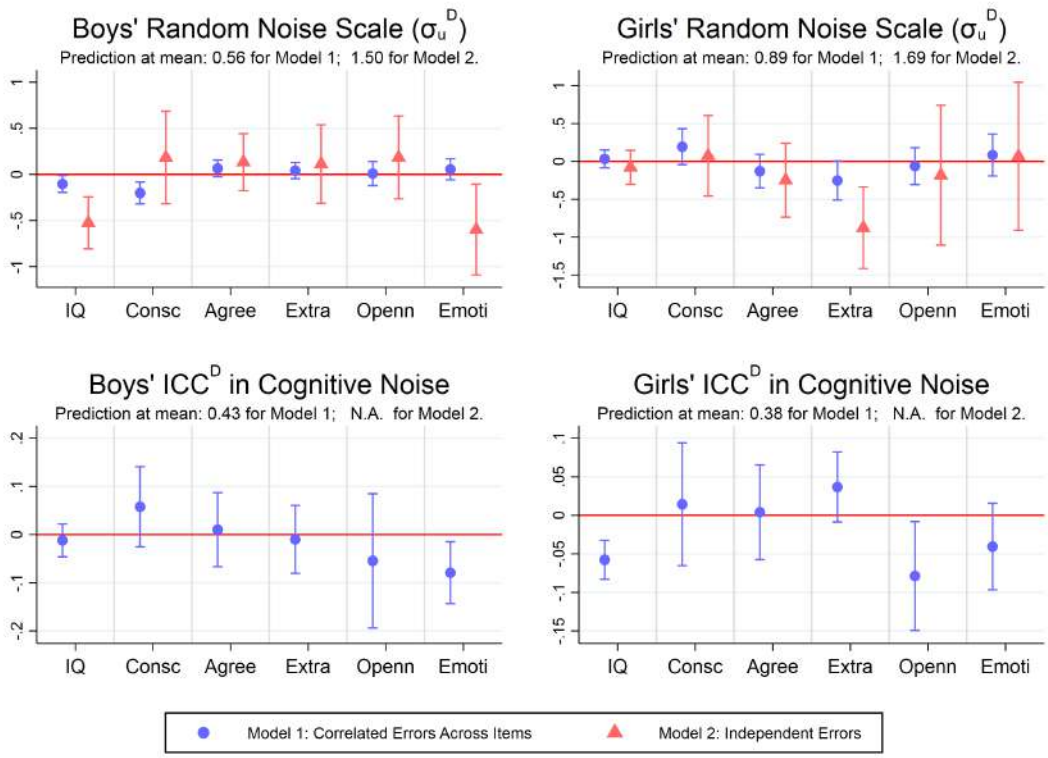


Figure F4: The Role of ICC: Decision Quality Parameters

Notes: This figure examines how excluding ICC affects decision quality parameters in the Expo-Power specification. Top panels show transitory noise σ_u^D, bottom panels show ICC^D. Left panels: boys, right panels: girls. Blue circles represent Model 1 with ICC, red diamonds represent Model 2 without ICC. Without ICC, transitory noise inflates by 169% for boys (σ_u^D: 0.56 vs. 1.50) and 90% for girls (0.89 vs. 1.69). Models without ICC cannot estimate ICC parameters (marked N.A.), forcing all variance into transitory noise and systematically misattributing persistent individual differences to random errors. This demonstrates that standard models severely overestimate apparent "irrationality." Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

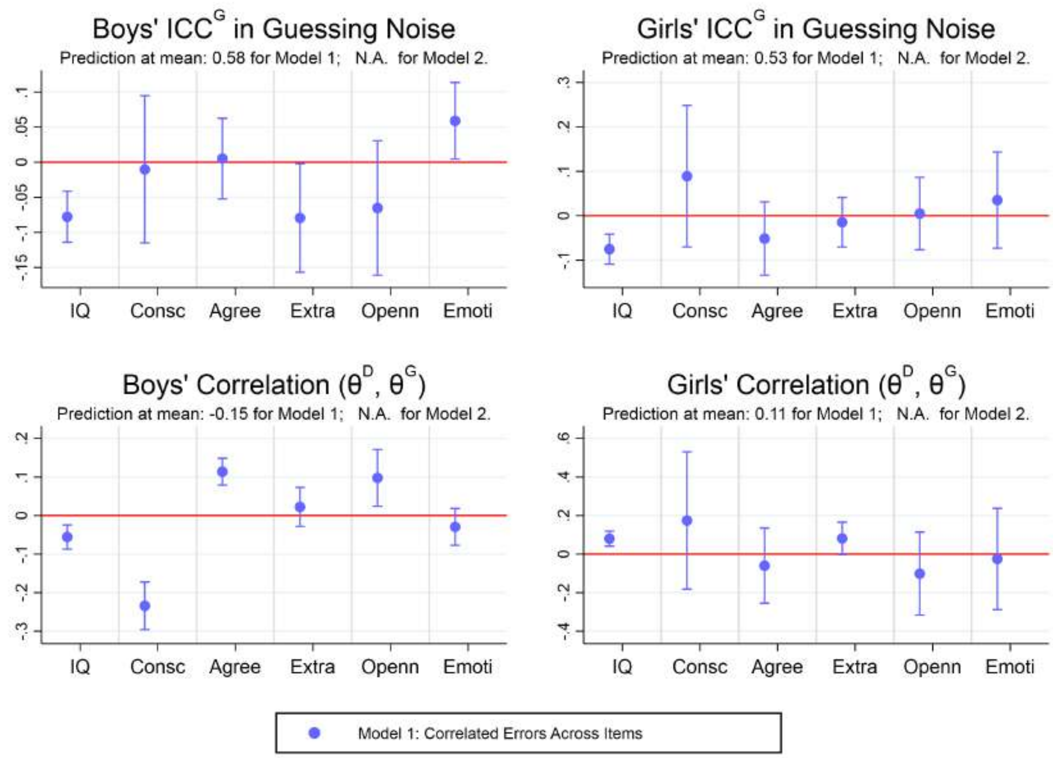


Figure F5: The Role of ICC: Guessing Structure Parameters

Notes: This figure shows ICC^G in guessing behavior and the correlation between deliberation and guessing random effects. Top panels show ICC^G, bottom panels show $\rho(\theta^D, \theta^G)$. Left panels: boys, right panels: girls. Blue circles represent Model 1 with ICC. Model 2 without ICC cannot estimate these parameters (marked N.A.). With ICC: $\text{ICC}^G = 0.58$ for boys, 0.53 for girls, $\rho = -0.15$ for boys, $+0.11$ for girls. These parameters reveal that strategic disengagement exhibits persistent trait-like structure and gender-specific correlation patterns that cannot be recovered without ICC modeling. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

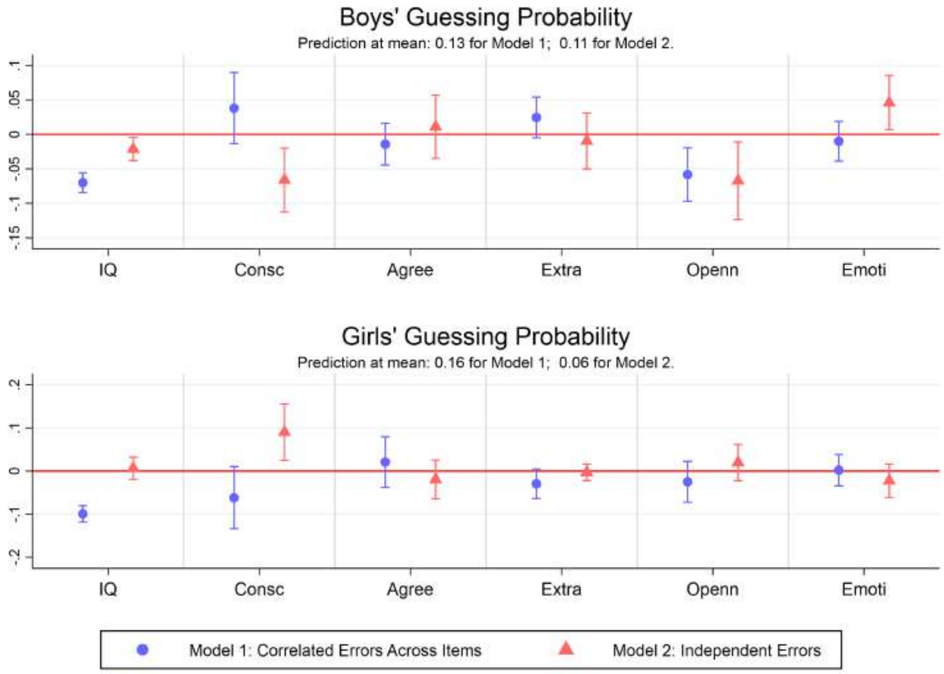


Figure F6: The Role of ICC: Guessing Probability

Notes: This figure shows how excluding ICC affects guessing probability. Top panel: boys, bottom panel: girls. Blue circles represent Model 1 with ICC, red diamonds represent Model 2 without ICC. Without ICC, guessing is underestimated: for boys, 0.13 with ICC vs. 0.11 without (15% underestimate), for girls, 0.16 with ICC vs. 0.06 without (63% underestimate). Models without ICC misallocate persistent strategic disengagement to transitory noise variance, incorrectly reducing estimated guessing rates. The 63% underestimation for girls is particularly severe, demonstrating that ICC structure is essential for accurately measuring attentional disengagement. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

F.2 CRRA Specification

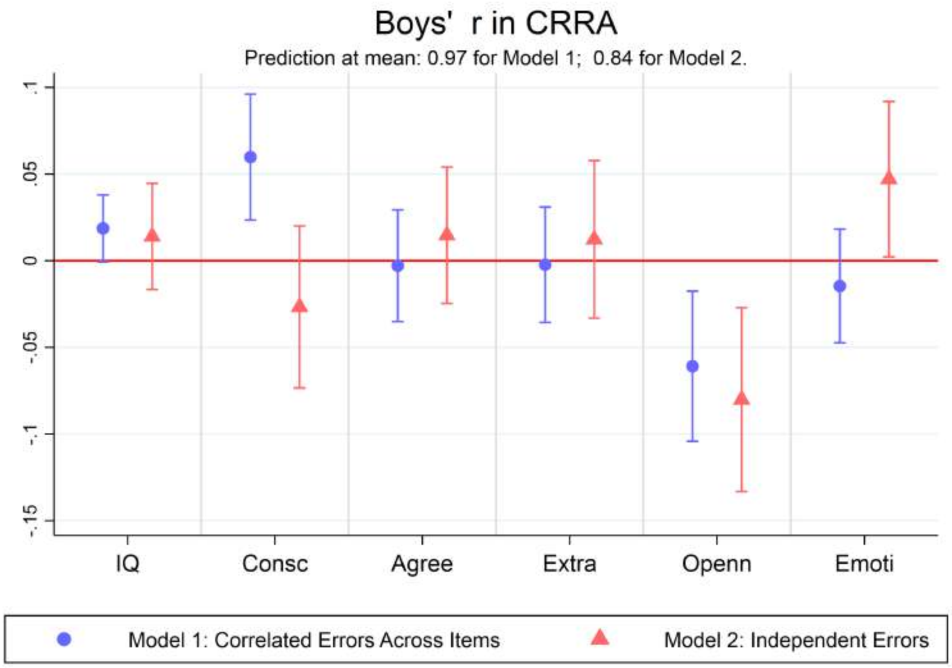


Figure F7: The Role of ICC: Risk Aversion in CRRA (Boys)

Notes: This figure shows how excluding ICC affects risk aversion r for boys in the CRRA specification. Blue circles represent Model 1 with ICC (correlated errors), red diamonds represent Model 2 without ICC (independent errors). Without ICC, risk aversion is underestimated: $r = 0.97$ with ICC vs. 0.84 without—a 13% underestimate. CRRA shows smaller bias than Expo-Power (13% vs. 19%), but the systematic underestimation confirms that ICC is essential regardless of functional form choice. Error bars show 90% confidence intervals. $N = 1{,}115$ boys.

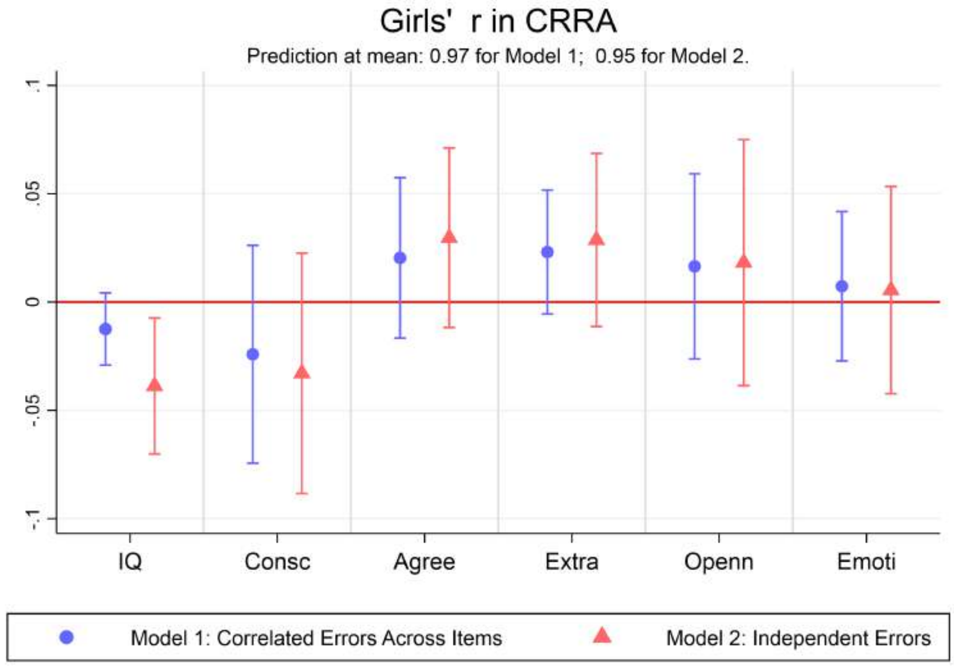


Figure F8: The Role of ICC: Risk Aversion in CRRA (Girls)

Notes: This figure shows how excluding ICC affects risk aversion r for girls in the CRRA specification. Blue circles represent Model 1 with ICC (correlated errors), red diamonds represent Model 2 without ICC (independent errors). Without ICC, risk aversion is underestimated: $r = 0.97$ with ICC vs. 0.95 without—a 2% underestimate. The small bias mirrors Expo-Power results (2% vs. 3%), indicating ICC omission affects girls' risk aversion estimates less than boys', though noise inflation remains severe. Error bars show 90% confidence intervals. $N = 1{,}089$ girls.

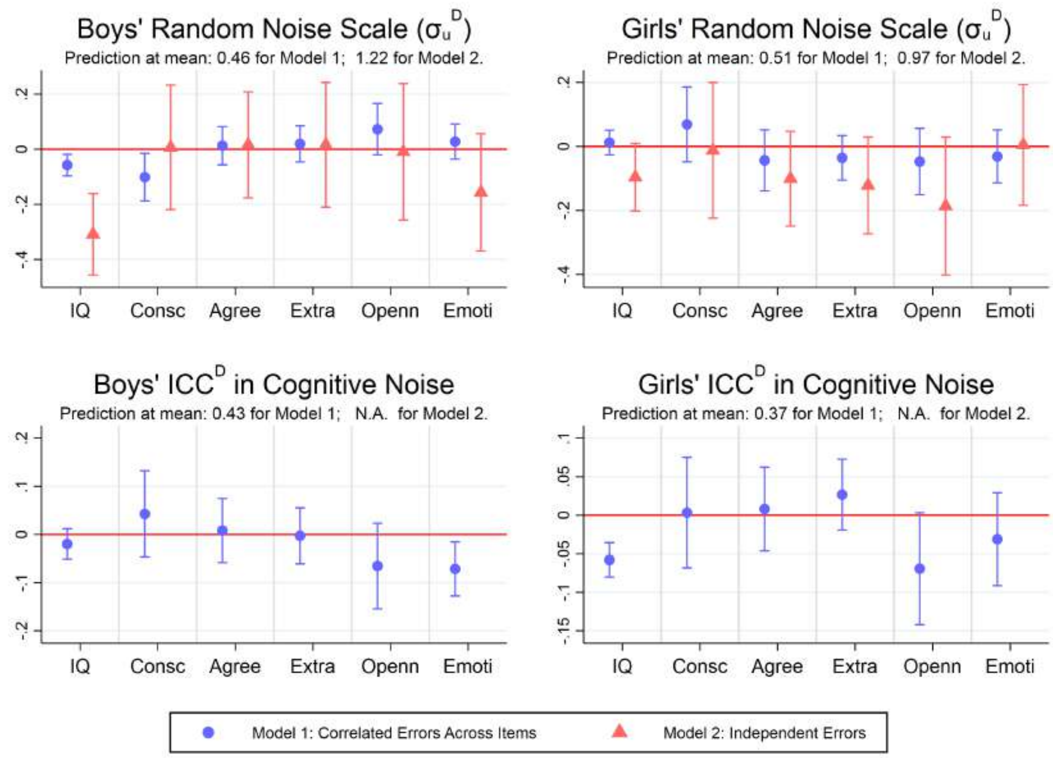


Figure F9: The Role of ICC: Decision Quality in CRRA

Notes: This figure examines how excluding ICC affects decision quality parameters in the CRRA specification. Top panels show transitory noise σ_u^D, bottom panels show ICC^D. Left panels: boys, right panels: girls. Blue circles represent Model 1 with ICC, red diamonds represent Model 2 without ICC. The pattern mirrors the Expo-Power specification: noise inflates by 165% for boys (σ_u^D: 0.46 vs. 1.22) and 90% for girls (0.51 vs. 0.97). ICC^D cannot be estimated without ICC structure (marked N.A.). The near-identical inflation rates across CRRA (90–165%) and Expo-Power (90–169%) confirm that ICC's role in separating persistent traits from transitory noise is independent of preference functional form. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

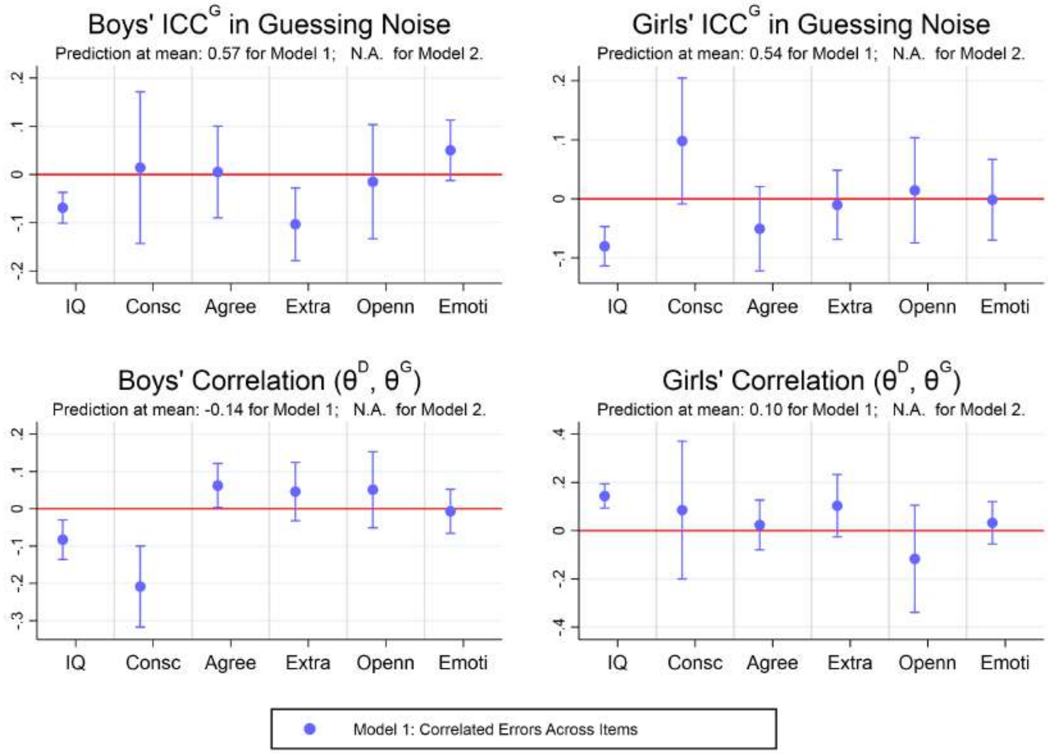


Figure F10: The Role of ICC: Guessing Structure in CRRA

Notes: This figure shows ICC^G in guessing behavior and the correlation between deliberation and guessing random effects in the CRRA specification. Top panels show ICC^G, bottom panels show $\rho(\theta^D, \theta^G)$. Left panels: boys, right panels: girls. Blue circles represent Model 1 with ICC. Model 2 cannot estimate these parameters (marked N.A.). With ICC: ICC^G = 0.57 for boys, 0.54 for girls, $\rho = -0.14$ for boys, $+0.10$ for girls. Values closely match Expo-Power estimates (ICC^G = 0.58 for boys, 0.53 for girls, $\rho = -0.15$ for boys, $+0.11$ for girls), demonstrating that guessing structure parameters are robust to functional form choice. Error bars show 90% confidence intervals. N = 1,115 boys, N = 1,089 girls.

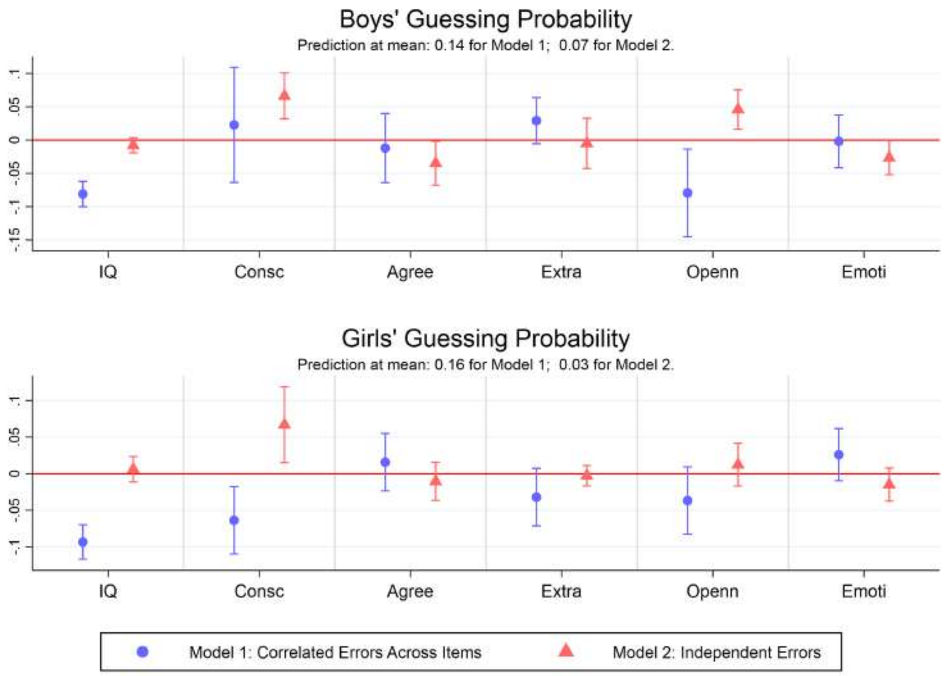


Figure F11: The Role of ICC: Guessing Probability in CRRA

Notes: This figure shows how excluding ICC affects guessing probability in the CRRA specification. Top panel: boys, bottom panel: girls. Blue circles represent Model 1 with ICC, red diamonds represent Model 2 without ICC. Without ICC, guessing is severely underestimated: for boys, 0.14 with ICC vs. 0.07 without (50% underestimate), for girls, 0.16 with ICC vs. 0.03 without (81% underestimate). The 81% underestimation for girls in CRRA exceeds even the 63% underestimation in Expo-Power, demonstrating that CRRA's functional form rigidity compounds ICC omission bias. Both specifications show systematic misallocation of persistent strategic disengagement to noise variance when ICC is excluded. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

The comparison between models with and without ICC reveals three critical insights. First, models without ICC overestimate transitory noise by 90–170%, conflating stable individual traits with random errors. This inflates apparent "irrationality" and obscures the fact that roughly 60% of decision inconsistency reflects persistent individual differences. Second, without ICC, strategic disengagement is misallocated to noise variance, underestimating true guessing rates by 15–63% in Expo-Power and 50–81% in CRRA. This masks the prevalence of attentional disengagement and its trait-like structure. Third, risk aversion parameters are systematically underestimated (up to 19% for boys in Expo-Power, 13% in CRRA), particularly the wealth-sensitivity component β (44% bias for boys). This demonstrates that failing to model persistent decision quality differences biases preference recovery itself, not just noise estimates.

These patterns explain why standard models in the literature often 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 3,200+ point log-likelihood improvements (Table 5) reflect genuine information about decision processes, not overfitting.

G Appendix G: Covariate Effects 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 2–6), 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. The comparison reveals that roughly 60% of apparent “noise” in standard models reflects stable individual differences rather than random decision errors—underscoring that ICC structure is essential for recovering trait-preference relationships, not merely a statistical refinement.

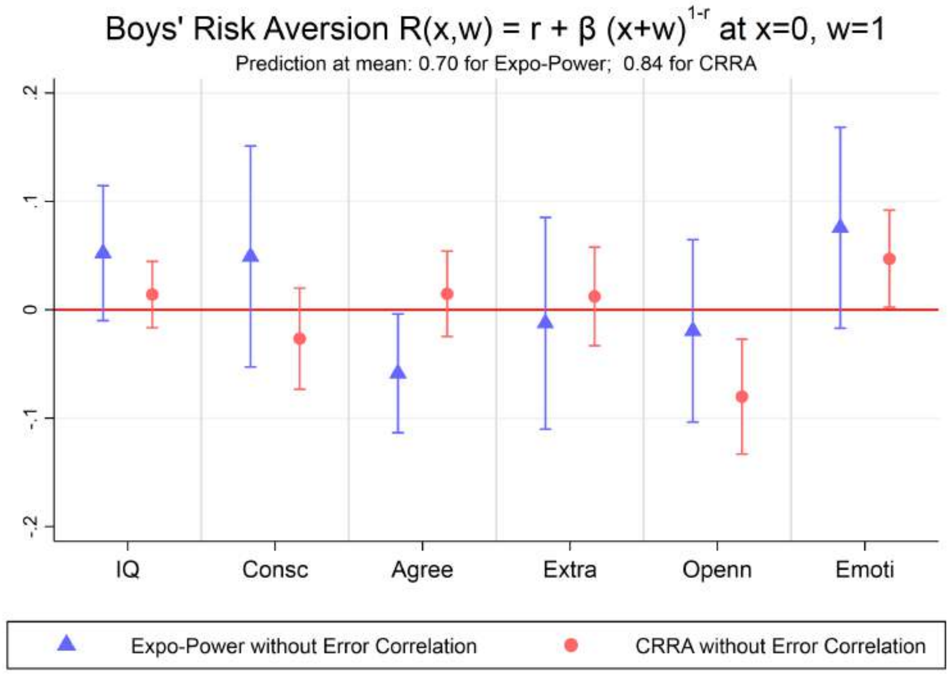


Figure G1: Covariate Effects on Risk Aversion Without ICC: Boys

Notes: This figure presents coefficient estimates showing how cognitive ability (IQ) and Big Five personality traits affect overall risk aversion $R(x, w) = r + \beta(x + w)^{1-r}$ evaluated at $x = 0$ and $w = 1$ for boys in models without ICC structure. Blue squares represent the Expo-Power specification, red circles represent the CRRA specification. Both specifications include guessing behavior but exclude ICC components. Compare to Figure 2: the prediction at the mean is 0.70 (Expo-Power) vs. 0.86 in the full model—a 19% underestimate. CRRA shows 0.84 vs. 0.97 in the full model—a 13% underestimate. Without ICC, models systematically underestimate risk aversion by conflating persistent individual precision differences with preference heterogeneity. Trait effects are substantially attenuated: the Conscientiousness coefficient drops from 0.120 ($p < 0.01$) in the full model to approximately 0.05 (not significant) without ICC. Error bars show 90% confidence intervals. $N = 1{,}115$ boys.

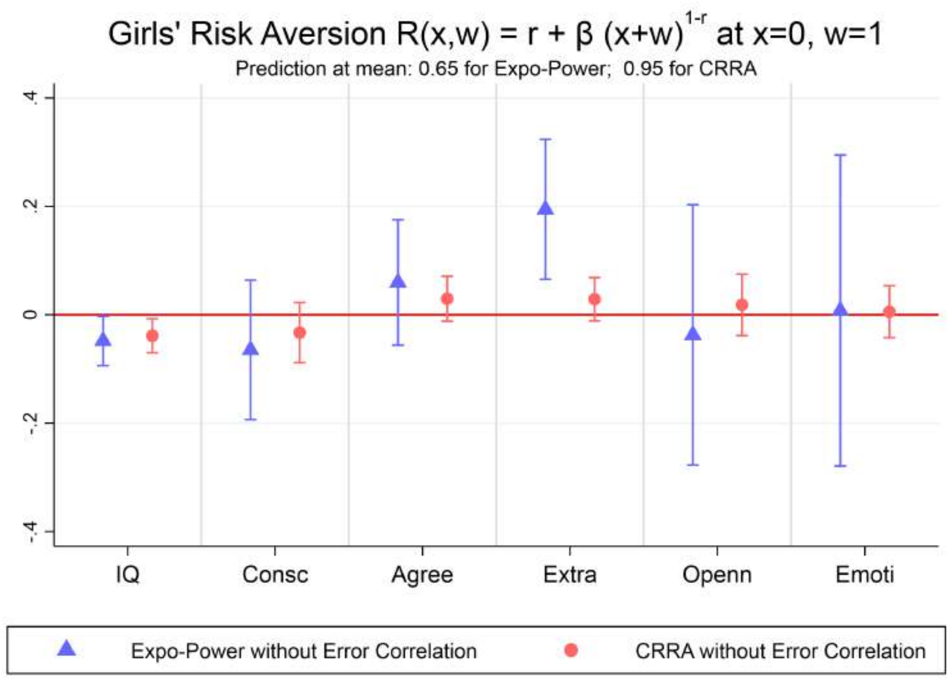


Figure G2: Covariate Effects on Risk Aversion Without ICC: Girls

Notes: This figure presents coefficient estimates showing how cognitive ability (IQ) and Big Five personality traits affect overall risk aversion $R(x, w) = r + \beta(x + w)^{1-r}$ evaluated at $x = 0$ and $w = 1$ for girls in models without ICC structure. Blue squares represent the Expo-Power specification, red circles represent the CRRA specification. Both specifications include guessing behavior but exclude ICC components. The prediction at the mean is 0.65 (Expo-Power) vs. 0.67 in the full model—relatively stable. CRRA shows 0.95 vs. 0.97 in the full model—also stable. However, the pattern of trait effects changes dramatically: Extraversion shows a much larger coefficient (approximately 0.19) with wider confidence intervals compared to 0.107 ($p < 0.05$) in the full model. The 32% gap between Expo-Power (0.65) and CRRA (0.95) persists even without ICC, confirming that functional form restrictions bind strongly for girls regardless of error structure specification. Error bars show 90% confidence intervals. $N = 1{,}089$ girls.

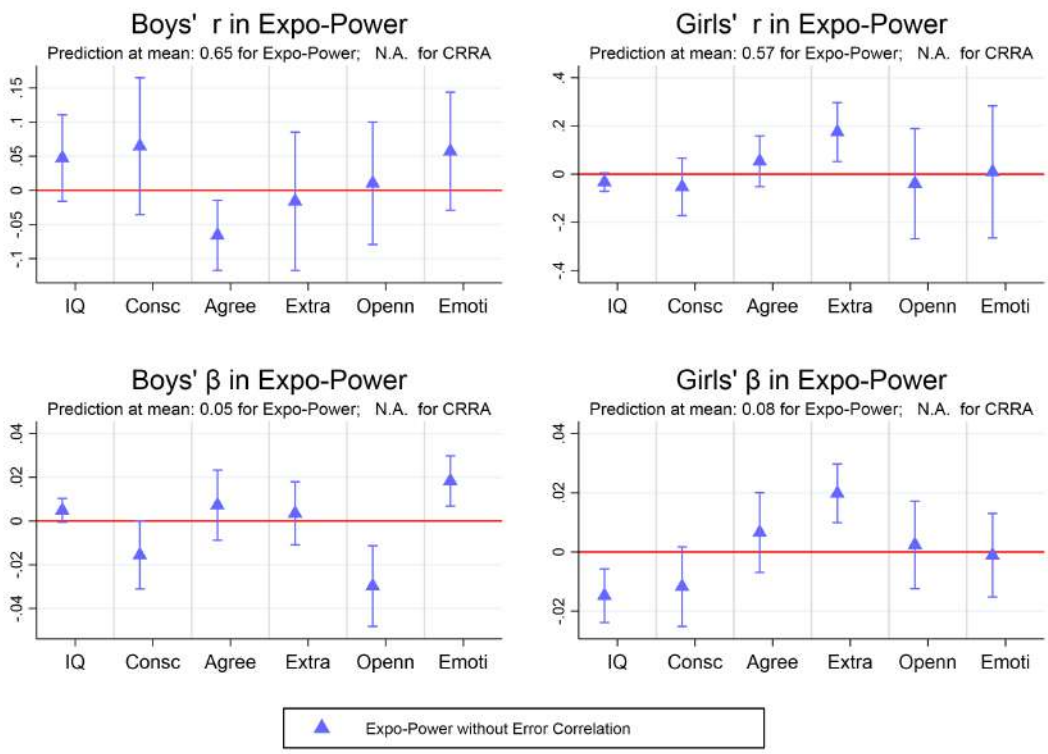


Figure G3: Decomposing Risk Aversion Without ICC: Covariate Effects on r and β Parameters

Notes: This figure decomposes the covariate effects on overall risk aversion $R(x,w) = r + \beta(x+w)^{1-r}$ into its two constituent parameters in the Expo-Power specification without ICC structure. Top panels show effects on baseline curvature parameter r, bottom panels show effects on wealth-sensitivity parameter β. Left panels show boys, right panels show girls. For boys (left column), the prediction at the mean is $r = 0.65$, $\beta = 0.05$. Compare to the full model (Figure 3): $r = 0.77$, $\beta = 0.09$. Both parameters are underestimated without ICC, with r showing 16% downward bias and β showing 44% downward bias. For girls (right column), the prediction at the mean is $r = 0.57$, $\beta = 0.08$. Compare to the full model: $r = 0.58$, $\beta = 0.09$. Girls show more stable estimates, though β still exhibits 11% downward bias. The pattern of larger bias in the wealth-sensitivity parameter β demonstrates that ICC structure is particularly important for recovering wealth-dependent risk aversion components. Coefficient patterns remain qualitatively similar but with substantially wider confidence intervals, indicating efficiency losses from misspecification. Error bars show 90% confidence intervals. CRRA cannot estimate separate r and β parameters (marked N.A.). $N = 1{,}115$ boys, $N = 1{,}089$ girls.

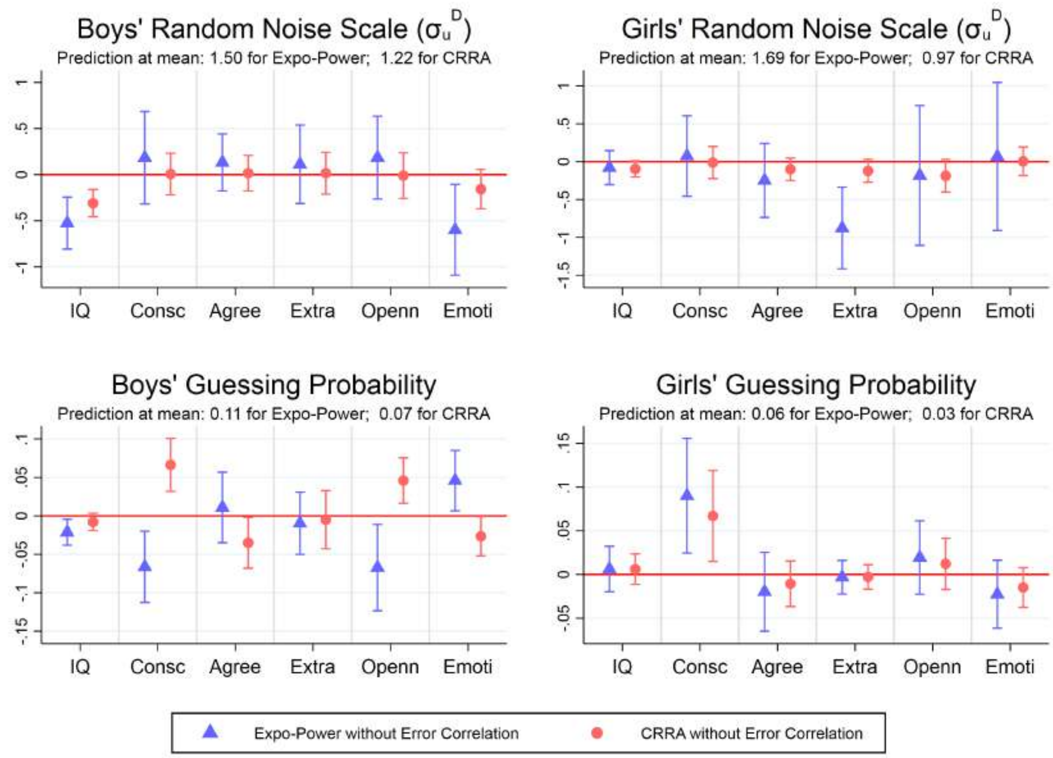


Figure G4: Decision Quality Without ICC: Random Noise and Guessing Probability

Notes: This figure examines how cognitive ability and personality traits affect decision quality parameters in models without ICC structure. Top panels show effects on the random noise scale σ_u^D, bottom panels show effects on guessing probability. Left panels show boys, right panels show girls. Blue squares represent Expo-Power, red circles represent CRRA. Critical finding: without ICC, noise estimates are severely inflated. For boys, $\sigma_u^D = 1.50$ (Expo-Power) vs. 0.56 in the full model with ICC (Figure 4)—a 169% inflation. For girls, $\sigma_u^D = 1.69$ (Expo-Power) vs. 0.89 in the full model—a 90% inflation. This demonstrates that standard models without ICC systematically misattribute persistent individual differences in decision precision to random noise, severely overestimating apparent "irrationality." Conversely, guessing probability is substantially underestimated: boys show 0.11 vs. 0.13 in the full model (15% underestimate), girls show 0.06 vs. 0.16 in the full model (63% underestimate). Models without ICC cannot distinguish persistent strategic disengagement from transitory noise, incorrectly allocating guessing behavior to noise variance. These systematic biases confirm that ICC structure is not a statistical refinement but a prerequisite for accurate preference and decision quality measurement. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

The comparison between models with and without ICC reveals three critical insights. First, noise inflation: models without ICC overestimate transitory noise by 90–170%, conflating stable individual traits with random errors. This inflates apparent "irrationality" and obscures the fact that roughly 60% of decision inconsistency reflects persistent individual differences. Second, guessing underestimation: without ICC, strategic disengagement is misallocated to noise variance, underestimating true guessing rates by 15–63%. This masks the prevalence of attentional disengagement and its trait-like structure. Third, preference attenuation: risk aversion parameters are systematically underestimated (up to 19% for boys), particularly the wealth-sensitivity component β (44% bias for boys). This demonstrates that failing to model persistent decision quality differences biases preference recovery itself, not just noise estimates.

These patterns explain why standard models in the literature often 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 3,200+ point log-likelihood improvements (Table 5) reflect genuine information about decision processes, not overfitting.

H Appendix H: Model Without Guessing

This appendix examines the role of explicit guessing behavior by comparing specifications that include strategic disengagement (baseline with ICC and guessing) to specifications that exclude guessing (ICC only). By removing the guessing component while retaining ICC structure, we isolate how misattributing strategic disengagement to deliberation noise affects parameter estimates. The comparison reveals that models without guessing inflate transitory noise by 90–120%, underestimate risk aversion by 4–9%, and substantially reduce ICC estimates. These systematic biases demonstrate that guessing behavior is not simply "more noise" but reflects a qualitatively distinct decision process—strategic attentional disengagement—that operates through separate mechanisms and must be explicitly modeled for accurate preference recovery.

H.1 Expo-Power Specification

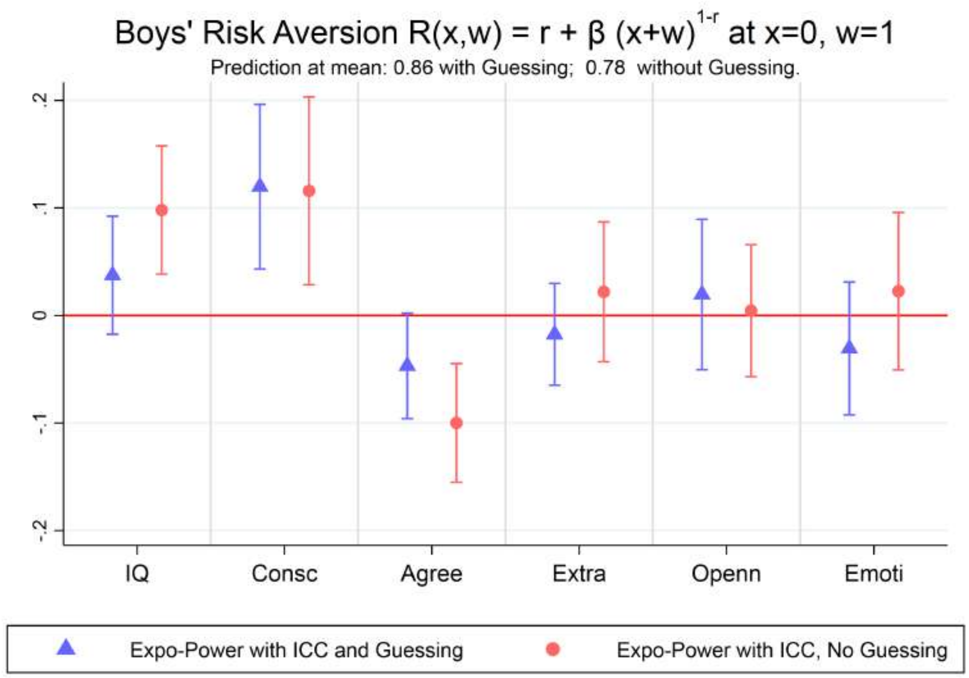


Figure H1: The Role of Guessing: Risk Aversion in Expo-Power (Boys)

Notes: This figure shows how excluding guessing behavior affects risk aversion $R(x, w) = r+\beta(x+w)^{1-r}$ evaluated at $x = 0$ and $w = 1$ for boys in the Expo-Power specification. Blue squares represent the baseline with guessing, red circles represent models without guessing. Without guessing, risk aversion is underestimated: $R(x, w) = 0.86$ with guessing vs. 0.78 without—a 9% underestimate. Error bars show 90% confidence intervals. $N = 1{,}115$ boys.

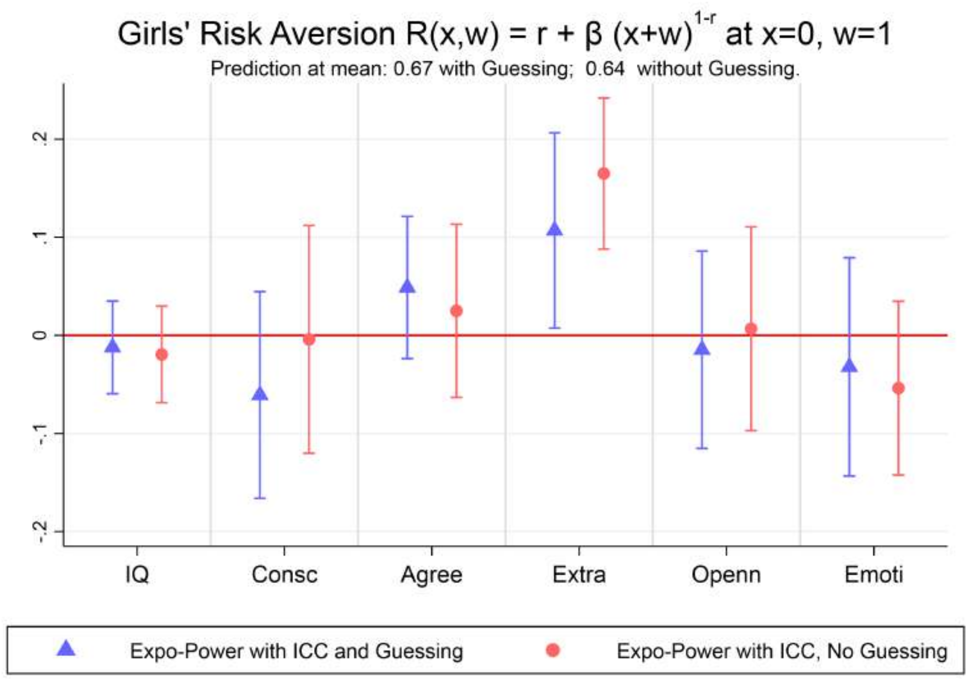


Figure H2: The Role of Guessing: Risk Aversion in Expo-Power (Girls)

Notes: This figure shows how excluding guessing behavior affects risk aversion $R(x, w) = r+\beta(x+w)^{1-r}$ evaluated at $x = 0$ and $w = 1$ for girls in the Expo-Power specification. Blue squares represent the baseline with guessing, red circles represent models without guessing. Without guessing, risk aversion is underestimated: $R(x, w) = 0.67$ with guessing vs. 0.64 without—a 4% underestimate. Error bars show 90% confidence intervals. $N = 1{,}089$ girls.

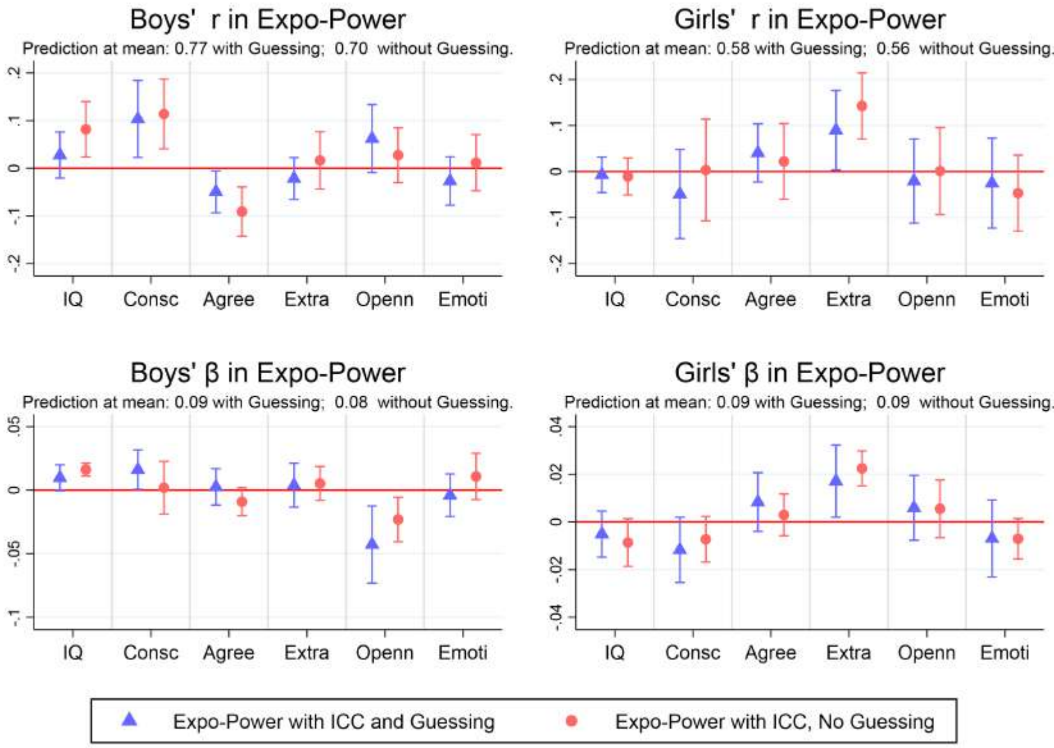


Figure H3: The Role of Guessing: Decomposition into r and β Parameters

Notes: This figure decomposes how excluding guessing affects risk aversion components in the Expo-Power specification. Top panels show baseline curvature r, bottom panels show wealth-sensitivity β. Left panels: boys, right panels: girls. Blue squares represent the baseline with guessing, red circles represent models without guessing. For boys: $r = 0.77$ with guessing vs. 0.70 without (9% underestimate), $\beta = 0.09$ vs. 0.08 (11% underestimate). For girls: $r = 0.58$ with guessing vs. 0.56 without (3% underestimate), $\beta = 0.09$ vs. 0.09 (stable). Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

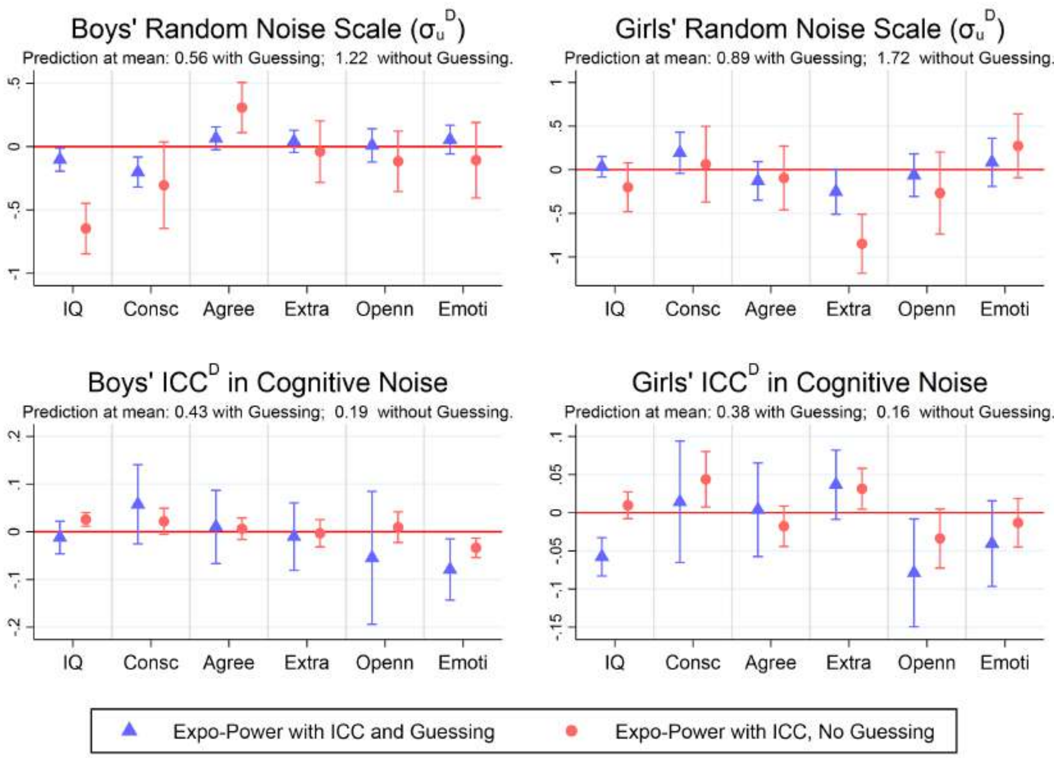


Figure H4: The Role of Guessing: Decision Quality in Expo-Power

Notes: This figure examines how excluding guessing affects decision quality parameters in the Expo-Power specification. Top panels show transitory noise σ_u^D, bottom panels show ICC^D. Left panels: boys, right panels: girls. Blue squares represent the baseline with guessing, red circles represent models without guessing. Without guessing, transitory noise inflates by 119% for boys (σ_u^D: 0.56 vs. 1.22) and 93% for girls (0.89 vs. 1.72). Conversely, ICC^D collapses by 56% for boys (0.43 vs. 0.19) and 58% for girls (0.38 vs. 0.16). This demonstrates that omitting guessing misallocates strategic disengagement to deliberation noise. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

H.2 CRRA Specification

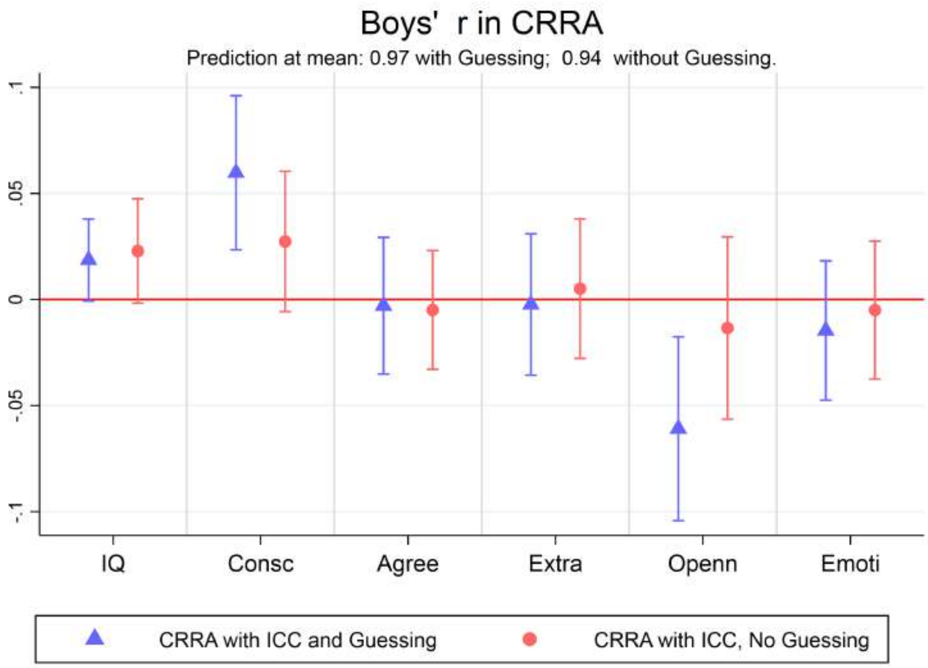


Figure H5: The Role of Guessing: Risk Aversion in CRRA (Boys)

Notes: This figure shows how excluding guessing affects risk aversion r for boys in the CRRA specification. Blue squares represent the baseline with guessing, red circles represent models without guessing. Without guessing, risk aversion is underestimated: $r = 0.97$ with guessing vs. 0.94 without—a 3% underestimate. CRRA shows smaller bias than Expo-Power (3% vs. 9%), reflecting functional form rigidity that constrains how guessing interacts with preferences. Error bars show 90% confidence intervals. $N = 1{,}115$ boys.

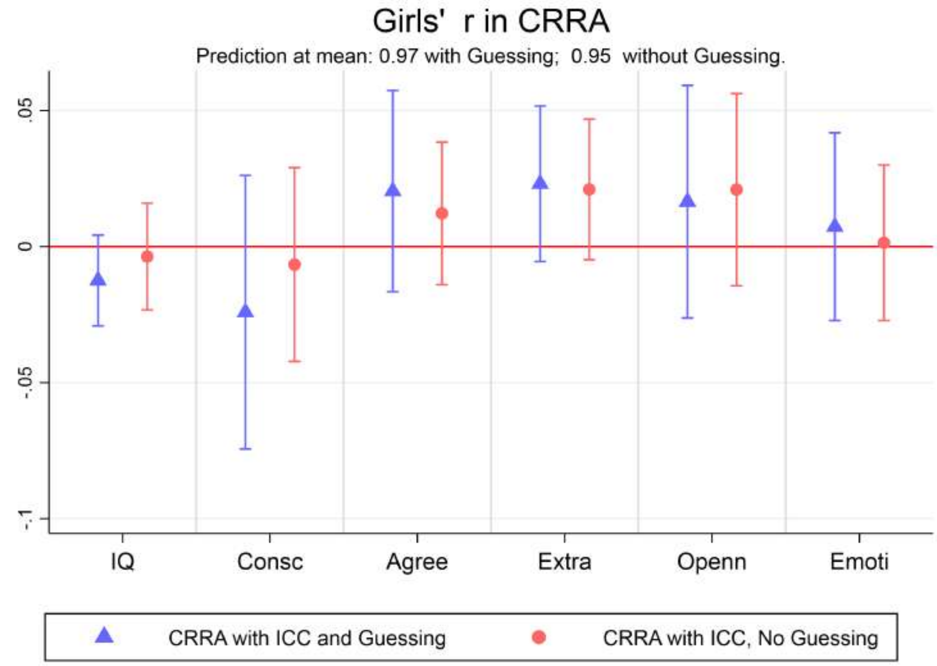


Figure H6: The Role of Guessing: Risk Aversion in CRRA (Girls)

Notes: This figure shows how excluding guessing affects risk aversion r for girls in the CRRA specification. Blue squares represent the baseline with guessing, red circles represent models without guessing. Without guessing, risk aversion is underestimated: $r = 0.97$ with guessing vs. 0.95 without—a 2% underestimate. The consistent 2–3% bias across both genders in CRRA contrasts with Expo-Power's larger gender differences (9% for boys, 4% for girls), indicating functional form constraints limit interaction channels. Error bars show 90% confidence intervals. $N = 1{,}089$ girls.

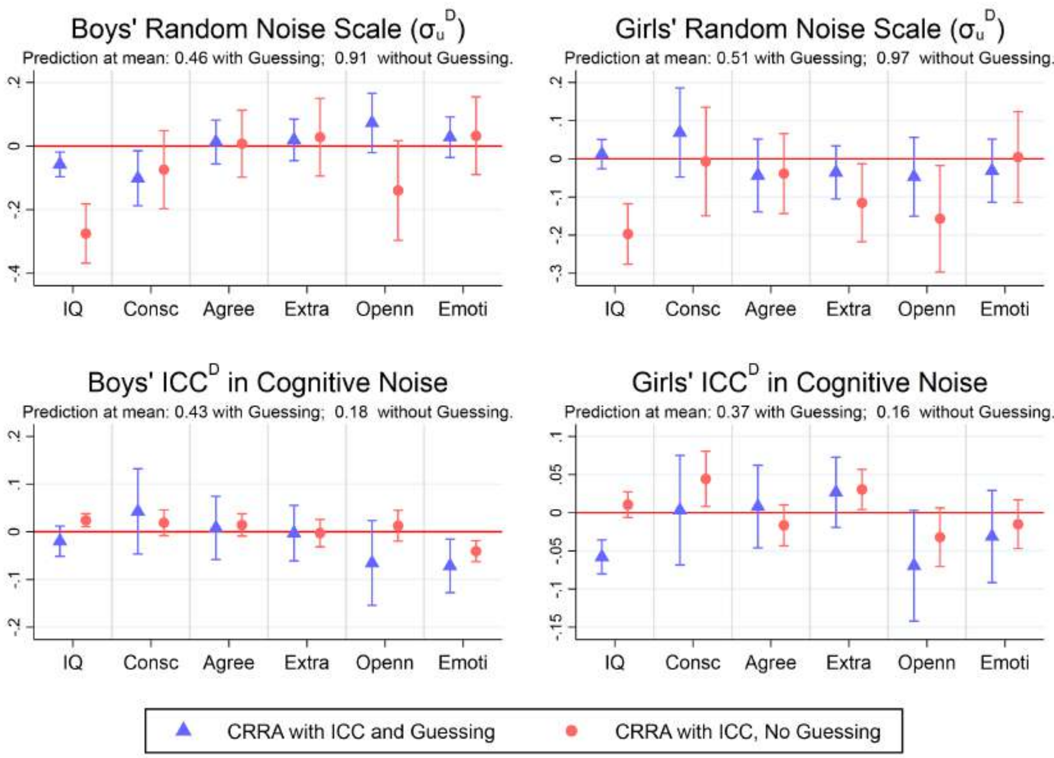


Figure H7: The Role of Guessing: Decision Quality in CRRA

Notes: This figure examines how excluding guessing affects decision quality parameters in the CRRA specification. Top panels show transitory noise σ_u^D, bottom panels show ICC^D. Left panels: boys, right panels: girls. Blue squares represent the baseline with guessing, red circles represent models without guessing. The pattern mirrors the Expo-Power specification: noise inflates by 98% for boys (σ_u^D: 0.46 vs. 0.91) and 90% for girls (0.51 vs. 0.97). ICC^D collapses by 58% for boys (0.43 vs. 0.18) and 57% for girls (0.37 vs. 0.16). Near-identical patterns across CRRA and Expo-Power confirm guessing captures genuine decision processes independent of preference functional form. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

I Appendix I: 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 both Expo-Power utility (Section I.1) and CRRA utility (Section I.2) to demonstrate that the patterns of how cognitive ability and personality traits affect guessing behavior are robust across utility specifications. In the coefficient plots, blue circles represent Model 1, red squares represent Model 2, and green triangles represent Model 3.

I.1 Expo-Power Specification

I.1.1 Parameter Estimates: Expo-Power Models

Table I1: Boys, Expo-Power, Model 1: Parameter Estimates

	Parameter								
	r in CRRA or $R(x, w)$ at $x=0$, $w=1$ in Expo-Power	r	β	Cognitive Random Noise s.d. (σ_{u^D})	ICC in Deliberation Equation	ICC in Guessing Equation	Correlation Between Deliberation and Guessing Noises	Probability of Guessing	Effect of σ_{u^D} on Guessing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	0.862***	0.770***	0.091***	0.556***	0.426***	0.580***	−0.151***	0.127***	
	(0.028)	(0.033)	(0.009)	(0.041)	(0.035)	(0.027)	(0.034)	(0.014)	
IQ	0.037	0.028	0.010	−0.103*	−0.012	−0.078***	−0.056***	−0.070***	
	(0.033)	(0.029)	(0.006)	(0.056)	(0.021)	(0.022)	(0.019)	(0.009)	
Conscientiousness	0.120**	0.104**	0.016*	−0.201***	0.058	−0.010	−0.234***	0.038	
	(0.047)	(0.049)	(0.009)	(0.072)	(0.051)	(0.064)	(0.038)	(0.031)	
Agreeableness	−0.047	−0.050*	0.003	0.065	0.010	0.005	0.114***	−0.014	
	(0.030)	(0.027)	(0.009)	(0.055)	(0.047)	(0.035)	(0.021)	(0.018)	
Extraversion	−0.018	−0.021	0.004	0.041	−0.010	−0.079*	0.022	0.025	
	(0.029)	(0.027)	(0.011)	(0.053)	(0.043)	(0.047)	(0.031)	(0.018)	
Openness	0.019	0.062	−0.043**	0.009	−0.055	−0.065	0.098**	−0.058**	
	(0.043)	(0.043)	(0.018)	(0.080)	(0.085)	(0.058)	(0.045)	(0.024)	
Emotional Stability	−0.031	−0.027	−0.004	0.056	−0.079**	0.059*	−0.029	−0.010	
	(0.038)	(0.031)	(0.010)	(0.069)	(0.039)	(0.033)	(0.029)	(0.018)	

Notes: Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Model 1 allows traits to affect guessing through stable individual differences.

Table I2: Boys, Expo-Power, Model 2: Parameter Estimates

	Parameter								
	r in CRRA or $R(x,w)$ at $x=0$, $w=1$ in Expo-Power	r	β	Cognitive Random Noise s.d. (σ_{u^D})	ICC in Deliberation Equation	ICC in Guessing Equation	Correlation Between Deliberation and Guessing Noises	Probability of Guessing	Effect of σ_{u^D} on Guessing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	0.863***	0.773***	0.090***	0.551***	0.425***	0.586***	−0.147***	0.124	0.932
	(0.027)	(0.031)	(0.009)	(0.039)	(0.029)	(0.025)	(0.039)	(0.318)	(0.615)
IQ	0.027	0.022	0.006	−0.077*	−0.025	−0.090***	−0.076***		
	(0.022)	(0.019)	(0.005)	(0.040)	(0.018)	(0.019)	(0.013)		
Conscientiousness	0.051*	0.043	0.008	−0.037	−0.022	0.099***	−0.328***		
	(0.028)	(0.028)	(0.010)	(0.034)	(0.037)	(0.035)	(0.054)		
Agreeableness	−0.025*	−0.033**	0.008	0.009	0.043	−0.024	0.145***		
	(0.014)	(0.016)	(0.007)	(0.012)	(0.033)	(0.025)	(0.034)		
Extraversion	−0.012	−0.018	0.005	0.028	−0.013	−0.097**	0.019		
	(0.017)	(0.019)	(0.008)	(0.020)	(0.031)	(0.039)	(0.042)		
Openness	0.052**	0.091***	−0.039***	−0.051**	−0.017	−0.058	0.086**		
	(0.021)	(0.024)	(0.013)	(0.022)	(0.057)	(0.049)	(0.040)		
Emotional Stability	−0.018	−0.017	−0.002	0.015	−0.054**	0.029	0.001		
	(0.017)	(0.017)	(0.007)	(0.017)	(0.026)	(0.030)	(0.064)		

Notes: Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Model 2 specifies that deliberation noise directly predicts guessing probability, and that IQ and personality traits affect guessing indirectly only through deliberation noise.

Table I3: Boys, Expo-Power, Model 3: Parameter Estimates

	Parameter								
	r in CRRA or $R(x,w)$ at $x=0$, $w=1$ in Expo-Power	r	β	Cognitive Random Noise s.d. (σ_{u^D})	ICC in Deliberation Equation	ICC in Guessing Equation	Correlation Between Deliberation and Guessing Noises	Probability of Guessing	Effect of σ_{u^D} on Guessing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	0.848***	0.754***	0.094***	0.573***	0.430***	0.574***	−0.156***	0.143	−0.409
	(0.028)	(0.032)	(0.008)	(0.045)	(0.051)	(0.031)	(0.042)	(0.168)	(0.291)
IQ	0.076***	0.061***	0.014***	−0.163***	−0.015	−0.042	−0.120***	−0.169***	
	(0.019)	(0.019)	(0.003)	(0.030)	(0.021)	(0.031)	(0.025)	(0.052)	
Conscientiousness	0.068	0.053	0.015	−0.117*	0.035	0.033	−0.239***	−0.034	
	(0.045)	(0.050)	(0.011)	(0.066)	(0.057)	(0.106)	(0.085)	(0.070)	
Agreeableness	−0.053**	−0.052*	−0.001	0.074*	0.013	−0.002	0.098***	0.022	
	(0.026)	(0.027)	(0.008)	(0.039)	(0.057)	(0.054)	(0.034)	(0.043)	
Extraversion	−0.009	−0.014	0.005	0.022	−0.003	−0.085	0.012	0.033	
	(0.030)	(0.031)	(0.010)	(0.055)	(0.047)	(0.080)	(0.110)	(0.032)	
Openness	−0.024	0.030	−0.054***	0.100	−0.074	−0.078	0.116	−0.017	
	(0.034)	(0.038)	(0.013)	(0.072)	(0.119)	(0.103)	(0.105)	(0.051)	
Emotional Stability	0.025	0.015	0.010	−0.051	−0.055	0.064	−0.021	−0.025	
	(0.027)	(0.024)	(0.008)	(0.051)	(0.043)	(0.047)	(0.032)	(0.038)	

Notes: Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Model 3 includes both direct trait effects and the deliberation noise channel.

Table I4: Girls, Expo-Power, Model 1: Parameter Estimates

	Parameter								
	r in CRRA or $R(x, w)$ at $x = 0$, $w = 1$ in Expo-Power	r	β	Cognitive Random Noise s.d. (σ_{u^D})	ICC in Deliberation Equation	ICC in Guessing Equation	Correlation Between Deliberation and Guessing Noises	Probability of Guessing	Effect of σ_{u^D} on Guessing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	0.669***	0.576***	0.093***	0.888***	0.377***	0.534***	0.115	0.160***	
	(0.028)	(0.025)	(0.005)	(0.071)	(0.019)	(0.025)	(0.074)	(0.017)	
IQ	−0.012	−0.007	−0.005	0.033	−0.058***	−0.075***	0.080***	−0.100***	
	(0.029)	(0.024)	(0.006)	(0.071)	(0.015)	(0.020)	(0.024)	(0.012)	
Conscientiousness	−0.061	−0.049	−0.012	0.194	0.014	0.089	0.174	−0.062	
	(0.064)	(0.059)	(0.008)	(0.144)	(0.048)	(0.097)	(0.216)	(0.044)	
Agreeableness	0.049	0.041	0.008	−0.129	0.004	−0.051	−0.060	0.021	
	(0.044)	(0.039)	(0.007)	(0.135)	(0.037)	(0.050)	(0.119)	(0.036)	
Extraversion	0.107*	0.090*	0.017*	−0.252	0.037	−0.015	0.081	−0.030	
	(0.061)	(0.052)	(0.009)	(0.157)	(0.028)	(0.034)	(0.051)	(0.021)	
Openness	−0.015	−0.021	0.006	−0.063	−0.079*	0.005	−0.101	−0.025	
	(0.061)	(0.056)	(0.008)	(0.148)	(0.043)	(0.050)	(0.131)	(0.029)	
Emotional Stability	−0.032	−0.025	−0.007	0.085	−0.041	0.035	−0.025	0.002	
	(0.068)	(0.059)	(0.010)	(0.168)	(0.034)	(0.066)	(0.160)	(0.022)	

Notes: Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Model 1 allows traits to affect guessing through stable individual differences.

Table I5: Girls, Expo-Power, Model 2: Parameter Estimates

	Parameter								
	r in CRRA or $R(x, w)$ at $x = 0$, $w = 1$ in Expo-Power	r	β	Cognitive Random Noise s.d. (σ_{u^D})	ICC in Deliberation Equation	ICC in Guessing Equation	Correlation Between Deliberation and Guessing Noises	Probability of Guessing	Effect of σ_{u^D} on Guessing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	0.674***	0.581***	0.094***	0.863***	0.383***	0.531***	0.108**	0.164	4.808***
	(0.027)	(0.023)	(0.004)	(0.063)	(0.019)	(0.023)	(0.050)	(0.996)	(1.163)
IQ	0.005	0.005	−0.000	−0.021***	−0.056***	−0.080***	0.085***		
	(0.011)	(0.011)	(0.002)	(0.005)	(0.013)	(0.017)	(0.016)		
Conscientiousness	0.001	−0.001	0.002	−0.010	0.035	0.079	0.143		
	(0.037)	(0.037)	(0.005)	(0.006)	(0.040)	(0.063)	(0.119)		
Agreeableness	0.010	0.011	−0.001	0.002	−0.009	−0.032	−0.041		
	(0.025)	(0.025)	(0.004)	(0.004)	(0.029)	(0.043)	(0.057)		
Extraversion	0.020	0.019	0.002	−0.008*	0.020	−0.004	0.118**		
	(0.020)	(0.021)	(0.003)	(0.004)	(0.027)	(0.031)	(0.055)		
Openness	−0.026	−0.027	0.001	−0.006	−0.074*	0.002	−0.065		
	(0.034)	(0.036)	(0.005)	(0.006)	(0.041)	(0.046)	(0.111)		
Emotional Stability	−0.010	−0.010	0.000	0.002	−0.040	0.024	−0.061		
	(0.026)	(0.027)	(0.004)	(0.004)	(0.034)	(0.043)	(0.060)		

Notes: Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Model 2 specifies that deliberation noise directly predicts guessing probability, and that IQ and personality traits affect guessing indirectly only through deliberation noise.

Table I6: Girls, Expo-Power, Model 3: Parameter Estimates

	Parameter								
	r in CRRA or $R(x,w)$ at $x=0$, $w=1$ in Expo-Power	r	β	Cognitive Random Noise s.d. (σ_{u^D})	ICC in Deliberation Equation	ICC in Guessing Equation	Correlation Between Deliberation and Guessing Noises	Probability of Guessing	Effect of σ_{u^D} on Guessing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	0.674***	0.580***	0.094***	0.868***	0.380***	0.528***	0.102**	0.181	−0.847
	(0.028)	(0.024)	(0.004)	(0.069)	(0.019)	(0.024)	(0.045)	(1.331)	(1.548)
IQ	−0.010	−0.006	−0.005	0.031	−0.060***	−0.075***	0.089***	−0.086	
	(0.037)	(0.029)	(0.008)	(0.096)	(0.015)	(0.019)	(0.017)	(0.054)	
Conscientiousness	−0.047	−0.039	−0.008	0.130	0.023	0.058	0.173	0.065	
	(0.051)	(0.047)	(0.009)	(0.110)	(0.043)	(0.068)	(0.129)	(0.136)	
Agreeableness	0.015	0.013	0.002	−0.024	−0.007	−0.046	−0.053	0.000	
	(0.039)	(0.034)	(0.008)	(0.095)	(0.030)	(0.051)	(0.066)	(0.066)	
Extraversion	0.103**	0.086*	0.016**	−0.244*	0.041	−0.019	0.130**	−0.248	
	(0.052)	(0.045)	(0.008)	(0.128)	(0.028)	(0.041)	(0.065)	(0.309)	
Openness	−0.031	−0.033	0.002	0.001	−0.085**	0.025	−0.065	−0.043	
	(0.062)	(0.054)	(0.011)	(0.165)	(0.042)	(0.051)	(0.127)	(0.152)	
Emotional Stability	−0.016	−0.012	−0.004	0.040	−0.040	0.038	−0.078	0.044	
	(0.052)	(0.045)	(0.009)	(0.129)	(0.033)	(0.047)	(0.073)	(0.079)	

Notes: Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Model 3 includes both direct trait effects and the deliberation noise channel.

I.1.2 Coefficient Plots: Expo-Power Models

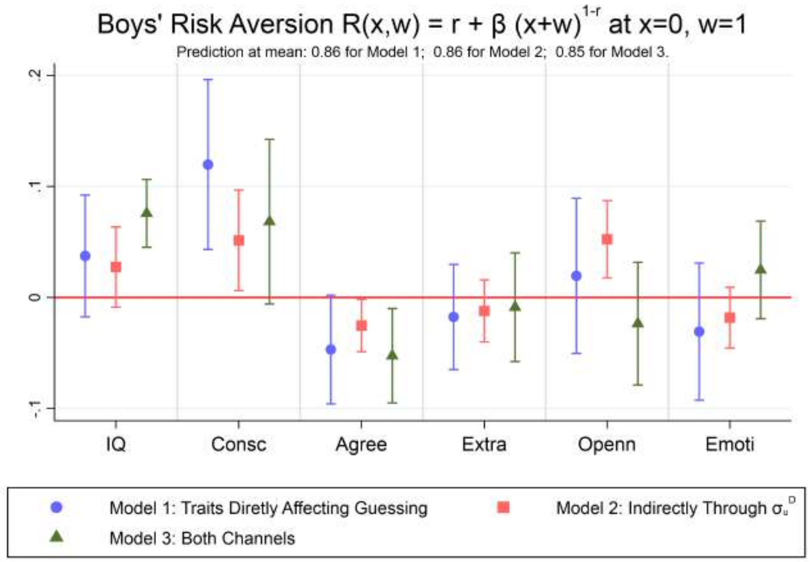


Figure I1: Risk Aversion Across Guessing Models: Boys

Notes: Effects of IQ or personality traits on risk aversion $R(x,w) = r + \beta(x+w)^{1-r}$ evaluated at $x = 0$ and $w = 1$ for boys. Blue circles represent Model 1 (traits affecting guessing), red squares represent Model 2 (σ_u^D affecting guessing), green triangles represent Model 3 (both σ_u^D and traits affect guessing). Predictions at mean: 0.86 for Model 1, 0.86 for Model 2, 0.85 for Model 3.

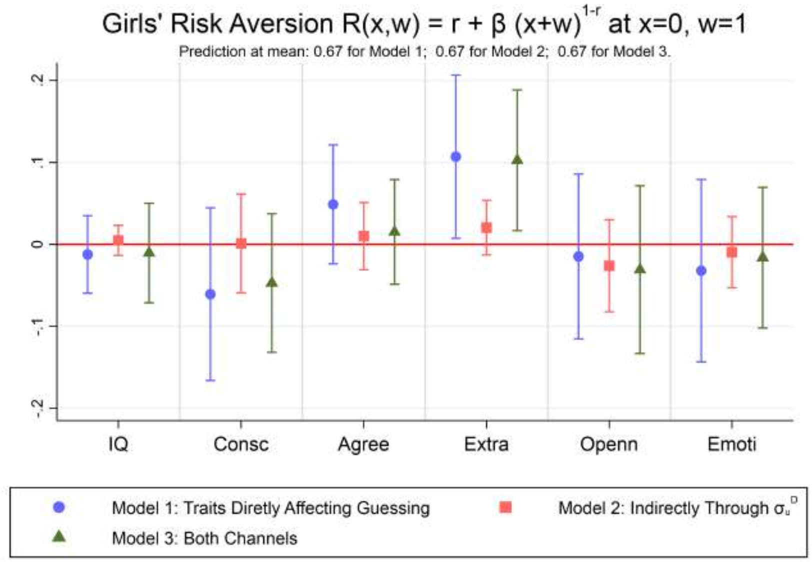


Figure I2: Risk Aversion Across Guessing Models: Girls

Notes: Effects of IQ or personality traits on risk aversion $R(x, w) = r + \beta(x + w)^{1-r}$ evaluated at $x = 0$ and $w = 1$ for girls. Blue circles represent Model 1 (traits affecting guessing), red squares represent Model 2 (σ_u^D affecting guessing), green triangles represent Model 3 (both σ_u^D and traits affect guessing). Predictions at mean: 0.67 for Model 1, 0.67 for Model 2, 0.67 for Model 3.

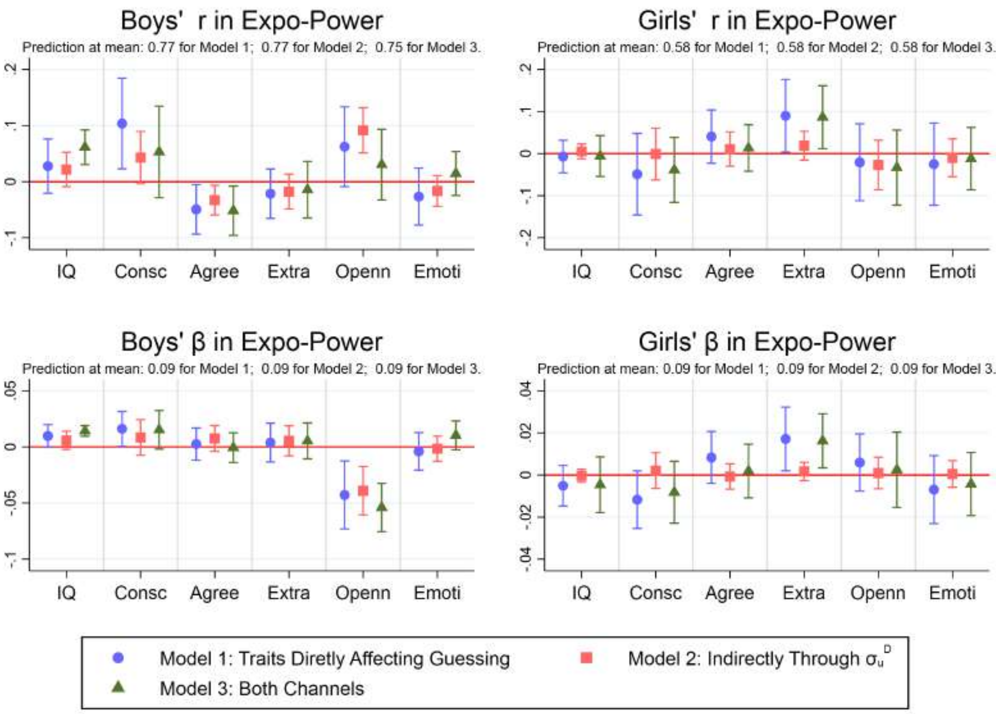


Figure I3: Expo-Power Parameters r and β Across Guessing Models

Notes: Top panels show effects of IQ or personality traits on parameter r for boys (left) and girls (right). Bottom panels show effects on parameter β for boys (left) and girls (right). Blue circles represent Model 1 (traits affecting guessing), red squares represent Model 2 (σ_u^D affecting guessing), green triangles represent Model 3 (both σ_u^D and traits affect guessing). Predictions at mean for r: boys = 0.77, 0.77, 0.75, girls = 0.58, 0.58, 0.58. For β: boys = 0.09, 0.09, 0.09, girls = 0.09, 0.09, 0.09.

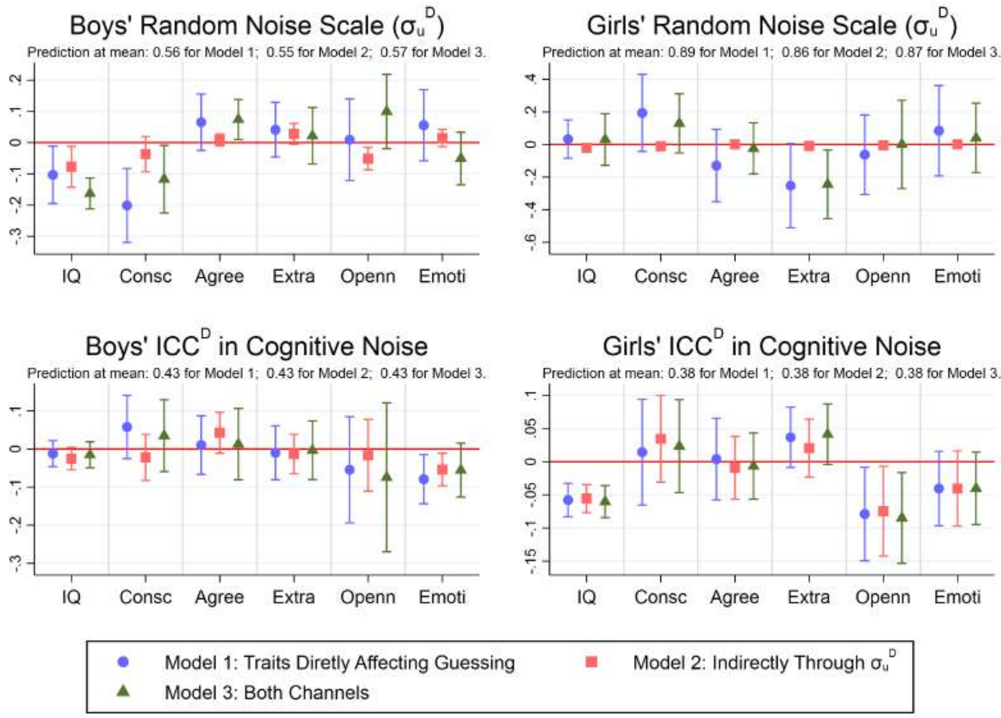


Figure I4: Random Noise Scale and ICC^D Across Guessing Models

Notes: Top panels show effects of IQ or personality traits on random noise scale (σ_u^D) for boys (left) and girls (right). Bottom panels show ICC^D for boys (left) and girls (right). Blue circles represent Model 1 (traits affecting guessing), red squares represent Model 2 (σ_u^D affecting guessing), green triangles represent Model 3 (both σ_u^D and traits affect guessing). Predictions at mean for σ_u^D: boys = 0.56, 0.55, 0.57, girls = 0.89, 0.88, 0.87. For ICC^D: boys = 0.43, 0.43, 0.43, girls = 0.38, 0.38, 0.38.

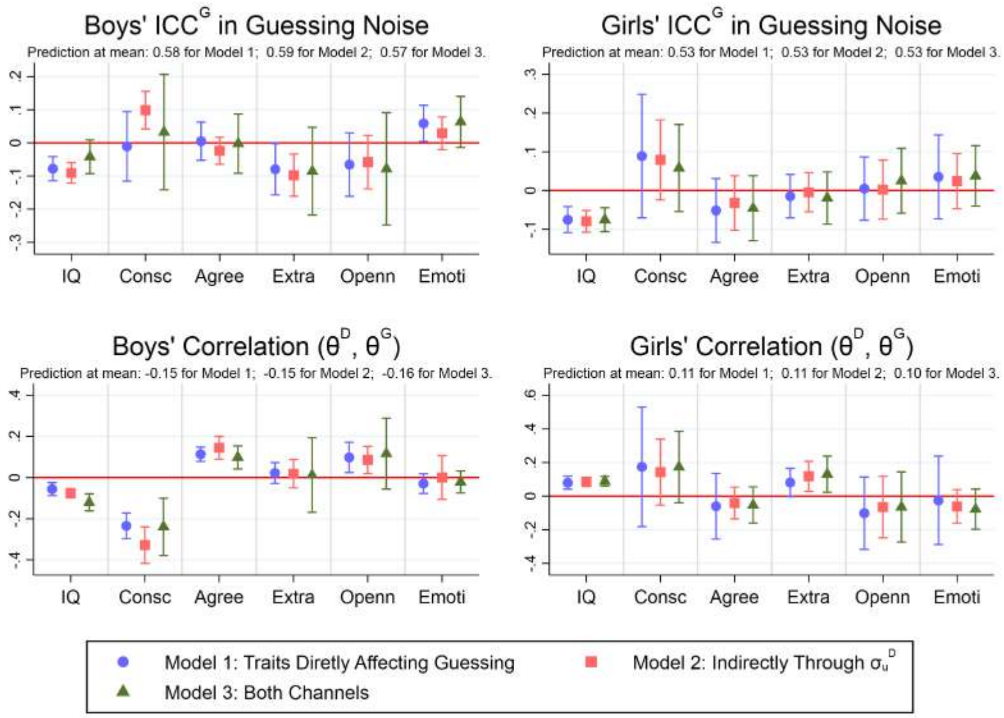


Figure I5: ICC^G and Correlation Between Random Effects Across Guessing Models

Notes: Top panels show ICC^G for boys (left) and girls (right). Bottom panels show correlation between deliberation and guessing random effects $\rho(\theta^D, \theta^G)$ for boys (left) and girls (right). Blue circles represent Model 1 (traits affecting guessing), red squares represent Model 2 (σ_u^D affecting guessing), green triangles represent Model 3 (both σ_u^D and traits affect guessing). Predictions at mean: ICC^G boys = 0.58, 0.59, 0.57 for Models 1–3, ICC^G girls = 0.53, 0.53, 0.53. Correlation boys = −0.15, −0.15, −0.16, girls = 0.11, 0.11, 0.10.

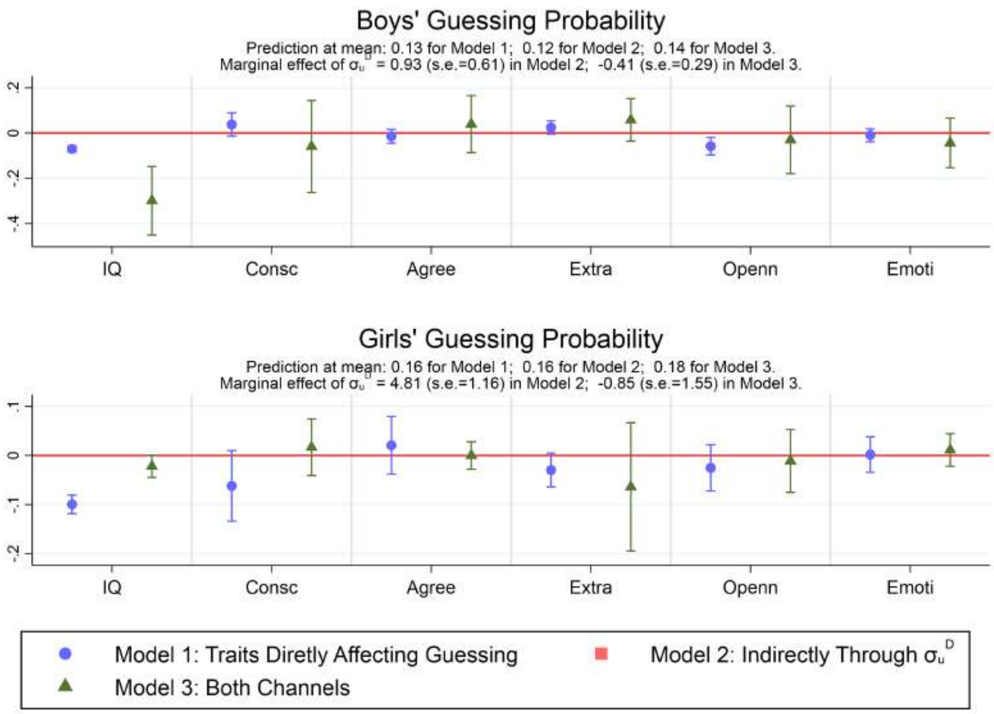


Figure I6: Guessing Probability Across Models

Notes: Effects of IQ or personality traits on guessing probability for boys (top) and girls (bottom). Blue circles represent Model 1 (traits affecting guessing), red squares represent Model 2 (σ_u^D affecting guessing), green triangles represent Model 3 (both σ_u^D and traits affect guessing). For Model 2, the marginal effect of σ_u^D is 0.93 (s.e. = 0.62) for boys and 4.80 (s.e. = 1.16) for girls. For Model 3, the marginal effect is -0.41 (s.e. = 0.29) for boys and -0.85 (s.e. = 1.55) for girls. Predictions at mean: boys = 0.13 for Model 1, 0.12 for Model 2, 0.14 for Model 3, girls = 0.16 for Model 1, 0.16 for Model 2, 0.18 for Model 3.

I.2 CRRA Specification

I.2.1 Parameter Estimates: CRRA Models

Table I7: Boys, CRRA, Model 1: Parameter Estimates

	Parameter								
	r in CRRA or $R(x, w)$ at $x = 0$, $w = 1$ in Expo-Power	r	β	Cognitive Random Noise s.d. (σ_{u^D})	ICC in Deliberation Equation	ICC in Guessing Equation	Correlation Between Deliberation and Guessing Noises	Probability of Guessing	Effect of σ_{u^D} on Guessing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant		0.970***		0.461***	0.432***	0.567***	−0.144***	0.137***	
		(0.012)		(0.026)	(0.025)	(0.024)	(0.036)	(0.015)	
IQ		0.019		−0.058**	−0.020	−0.069***	−0.083**	−0.081***	
		(0.012)		(0.024)	(0.019)	(0.019)	(0.032)	(0.012)	
Conscientiousness		0.060***		−0.101*	0.043	0.014	−0.209***	0.023	
		(0.022)		(0.053)	(0.054)	(0.096)	(0.066)	(0.053)	
Agreeableness		−0.003		0.013	0.008	0.005	0.062*	−0.012	
		(0.020)		(0.042)	(0.040)	(0.058)	(0.036)	(0.032)	
Extraversion		−0.002		0.019	−0.003	−0.103**	0.046	0.029	
		(0.020)		(0.040)	(0.035)	(0.046)	(0.047)	(0.021)	
Openness		−0.061**		0.073	−0.065	−0.015	0.051	−0.079**	
		(0.026)		(0.057)	(0.054)	(0.072)	(0.062)	(0.040)	
Emotional Stability		−0.015		0.028	−0.071**	0.050	−0.007	−0.002	
		(0.020)		(0.039)	(0.034)	(0.038)	(0.036)	(0.024)	

Notes: Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Model 1 allows traits to affect guessing through stable individual differences.

Table I8: Boys, CRRA, Model 2: Parameter Estimates

	Parameter								
	r in CRRA or $R(x, w)$ at $x = 0$, $w = 1$ in Expo-Power	r	β	Cognitive Random Noise s.d. (σ_{u^D})	ICC in Deliberation Equation	ICC in Guessing Equation	Correlation Between Deliberation and Guessing Noises	Probability of Guessing	Effect of σ_{u^D} on Guessing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant		0.968^{***}		0.466^{***}	0.425^{***}	0.571^{***}	-0.144^{***}	0.132	2.498
		(0.012)		(0.024)	(0.028)	(0.025)	(0.033)	(0.777)	(1.722)
IQ		0.005		-0.033	-0.033^{**}	-0.077^{***}	-0.082^{***}		
		(0.011)		(0.020)	(0.016)	(0.017)	(0.021)		
Conscientiousness		0.016		-0.006	-0.007	0.052	-0.219^{***}		
		(0.014)		(0.010)	(0.060)	(0.040)	(0.047)		
Agreeableness		0.004		-0.002	0.018	-0.002	0.064^{**}		
		(0.010)		(0.006)	(0.050)	(0.028)	(0.032)		
Extraversion		0.002		0.012	0.000	-0.109^{***}	0.040		
		(0.011)		(0.010)	(0.039)	(0.038)	(0.035)		
Openness		-0.014		-0.022^{*}	-0.012	-0.025	0.043		
		(0.016)		(0.012)	(0.086)	(0.049)	(0.050)		
Emotional Stability		-0.004		0.003	-0.055^{**}	0.037	-0.000		
		(0.010)		(0.006)	(0.027)	(0.034)	(0.029)		

Notes: Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Model 2 specifies that deliberation noise directly predicts guessing probability, and that IQ and personality traits affect guessing indirectly only through deliberation noise.

Table I9: Boys, CRRA, Model 3: Parameter Estimates

	Parameter								
	r in CRRA or $R(x, w)$ at $x = 0$, $w = 1$ in Expo-Power	r	β	Cognitive Random Noise s.d. (σ_{u^D})	ICC in Deliberation Equation	ICC in Guessing Equation	Correlation Between Deliberation and Guessing Noises	Probability of Guessing	Effect of σ_{u^D} on Guessing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant		0.971^{***}		0.460^{***}	0.428^{***}	0.566^{***}	-0.149^{***}	0.149	-3.328^{*}
		(0.012)		(0.024)	(0.027)	(0.026)	(0.051)	(0.890)	(1.950)
IQ		0.017^{*}		-0.053^{***}	-0.024	-0.055^{*}	-0.046^{*}	-0.271^{***}	
		(0.010)		(0.019)	(0.020)	(0.029)	(0.026)	(0.071)	
Conscientiousness		0.048^{***}		-0.076^{**}	0.033	0.030	-0.240^{***}	-0.244^{**}	
		(0.018)		(0.031)	(0.047)	(0.066)	(0.061)	(0.119)	
Agreeableness		-0.007		0.016	0.008	-0.012	0.136^{***}	0.056	
		(0.014)		(0.026)	(0.038)	(0.036)	(0.041)	(0.083)	
Extraversion		0.013		-0.012	0.011	-0.118^{**}	0.057	0.001	
		(0.016)		(0.027)	(0.038)	(0.056)	(0.041)	(0.096)	
Openness		-0.073^{***}		0.092^{***}	-0.063	-0.044	0.029	0.235	
		(0.020)		(0.034)	(0.060)	(0.081)	(0.058)	(0.155)	
Emotional Stability		0.008		-0.012	-0.056^{*}	0.064	-0.027	-0.045	
		(0.017)		(0.028)	(0.033)	(0.043)	(0.031)	(0.104)	

Notes: Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Model 3 includes both direct trait effects and the deliberation noise channel.

Table I10: Girls, CRRA, Model 1: Parameter Estimates

	Parameter								
	r in CRRA or $R(x,w)$ at $x=0$, $w=1$ in Expo-Power	r	β	Cognitive Random Noise s.d. (σ_{u^D})	ICC in Deliberation Equation	ICC in Guessing Equation	Correlation Between Deliberation and Guessing Noises	Probability of Guessing	Effect of σ_{u^D} on Guessing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant		0.970***		0.509***	0.375***	0.536***	0.101*	0.162***	
		(0.011)		(0.026)	(0.019)	(0.023)	(0.057)	(0.014)	
IQ		−0.012		0.012	−0.058***	−0.080***	0.144***	−0.093***	
		(0.010)		(0.023)	(0.014)	(0.020)	(0.031)	(0.014)	
Conscientiousness		−0.024		0.069	0.003	0.098	0.086	−0.064**	
		(0.031)		(0.071)	(0.044)	(0.065)	(0.174)	(0.028)	
Agreeableness		0.020		−0.044	0.008	−0.051	0.024	0.016	
		(0.023)		(0.058)	(0.033)	(0.043)	(0.063)	(0.024)	
Extraversion		0.023		−0.036	0.027	−0.010	0.103	−0.032	
		(0.017)		(0.042)	(0.028)	(0.036)	(0.079)	(0.024)	
Openness		0.016		−0.047	−0.069	0.014	−0.117	−0.037	
		(0.026)		(0.063)	(0.044)	(0.054)	(0.135)	(0.028)	
Emotional Stability		0.007		−0.031	−0.031	−0.001	0.033	0.026	
		(0.021)		(0.050)	(0.037)	(0.042)	(0.053)	(0.022)	

Notes: Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Model 1 allows traits to affect guessing through stable individual differences.

Table I11: Girls, CRRA, Model 2: Parameter Estimates

	Parameter								
	r in CRRA or $R(x,w)$ at $x=0$, $w=1$ in Expo-Power	r	β	Cognitive Random Noise s.d. (σ_{u^D})	ICC in Deliberation Equation	ICC in Guessing Equation	Correlation Between Deliberation and Guessing Noises	Probability of Guessing	Effect of σ_{u^D} on Guessing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant		0.972***		0.504***	0.377***	0.539***	0.111**	0.163	14.199***
		(0.010)		(0.024)	(0.019)	(0.023)	(0.054)	(0.432)	(0.915)
IQ		−0.002		−0.007***	−0.055***	−0.081***	0.079***		
		(0.005)		(0.001)	(0.011)	(0.017)	(0.015)		
Conscientiousness		0.008		−0.004**	0.018	0.090	0.130		
		(0.013)		(0.002)	(0.037)	(0.058)	(0.137)		
Agreeableness		0.003		0.001	−0.005	−0.037	−0.039		
		(0.010)		(0.001)	(0.025)	(0.038)	(0.055)		
Extraversion		0.009		−0.003**	0.020	0.003	0.099		
		(0.007)		(0.001)	(0.026)	(0.030)	(0.072)		
Openness		−0.006		−0.002	−0.071*	−0.004	−0.037		
		(0.010)		(0.002)	(0.041)	(0.043)	(0.142)		
Emotional Stability		−0.003		0.001	−0.039	0.016	−0.048		
		(0.010)		(0.001)	(0.036)	(0.041)	(0.061)		

Notes: Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Model 2 specifies that deliberation noise directly predicts guessing probability, and that IQ and personality traits affect guessing indirectly only through deliberation noise.

Table I12: Girls, CRRA, Model 3: Parameter Estimates

	Parameter								
	r in CRRA or $R(x, w)$ at $x = 0$, $w = 1$ in Expo-Power	r	β	Cognitive Random Noise s.d. (σ_{u^D})	ICC in Deliberation Equation	ICC in Guessing Equation	Correlation Between Deliberation and Guessing Noises	Probability of Guessing	Effect of σ_{u^D} on Guessing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant		0.971***		0.508***	0.374***	0.539***	0.108**	0.176	−3.819
		(0.011)		(0.027)	(0.019)	(0.025)	(0.043)	(3.472)	(6.890)
IQ		−0.010		0.012	−0.059***	−0.077***	0.076***	−0.066	
		(0.016)		(0.033)	(0.013)	(0.020)	(0.021)	(0.082)	
Conscientiousness		−0.020		0.058	0.006	0.092	0.166	0.155	
		(0.021)		(0.040)	(0.040)	(0.085)	(0.152)	(0.300)	
Agreeableness		0.009		−0.019	0.002	−0.064	−0.040	−0.048	
		(0.020)		(0.039)	(0.026)	(0.051)	(0.122)	(0.091)	
Extraversion		0.032		−0.058	0.035	−0.012	0.093	−0.259	
		(0.023)		(0.052)	(0.029)	(0.044)	(0.069)	(0.545)	
Openness		0.013		−0.043	−0.072	−0.003	−0.103	−0.198	
		(0.048)		(0.114)	(0.047)	(0.072)	(0.091)	(0.248)	
Emotional Stability		−0.006		0.006	−0.039	0.034	−0.049	0.030	
		(0.017)		(0.040)	(0.037)	(0.051)	(0.071)	(0.165)	

Notes: Standard errors in parentheses are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Model 3 includes both direct trait effects and the deliberation noise channel.

I.2.2 Coefficient Plots: CRRA Models

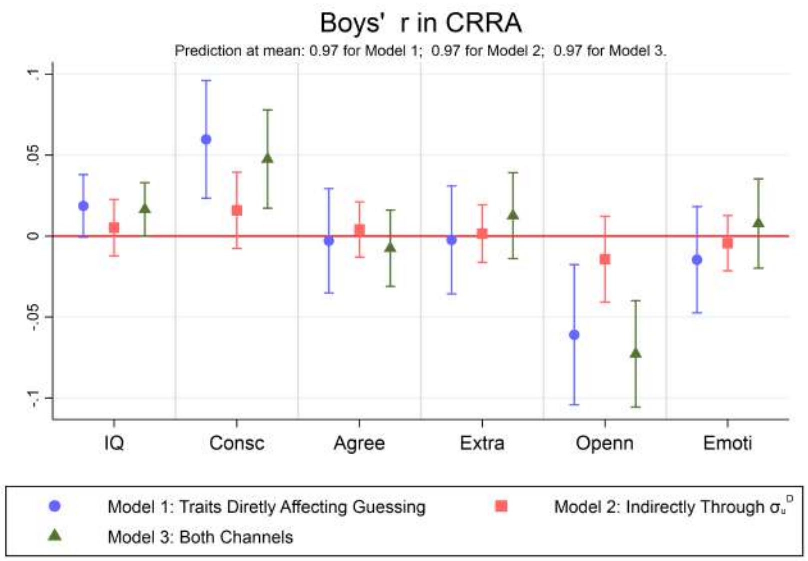


Figure I7: Risk Aversion Parameter r Across Guessing Models: Boys (CRRA)

Notes: Effects of IQ or personality traits on risk aversion parameter r in CRRA utility for boys. Blue circles represent Model 1 (traits affecting guessing), red squares represent Model 2 (σ_u^D affecting guessing), green triangles represent Model 3 (both σ_u^D and traits affect guessing). Predictions at mean: 0.97 for Model 1, 0.97 for Model 2, 0.97 for Model 3.

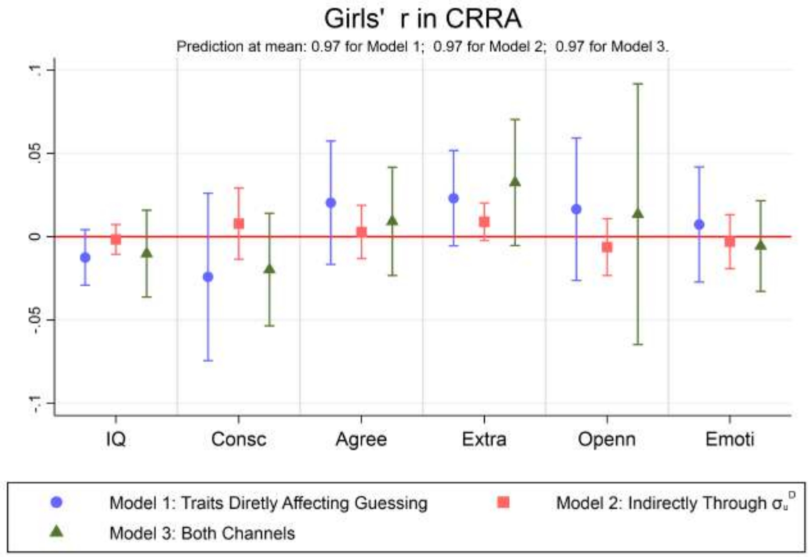


Figure I8: Risk Aversion Parameter r Across Guessing Models: Girls (CRRA)

Notes: Effects of IQ or personality traits on risk aversion parameter r in CRRA utility for girls. Blue circles represent Model 1 (traits affecting guessing), red squares represent Model 2 (σ_u^D affecting guessing), green triangles represent Model 3 (both σ_u^D and traits affect guessing). Predictions at mean: 0.97 for Model 1, 0.97 for Model 2, 0.97 for Model 3.

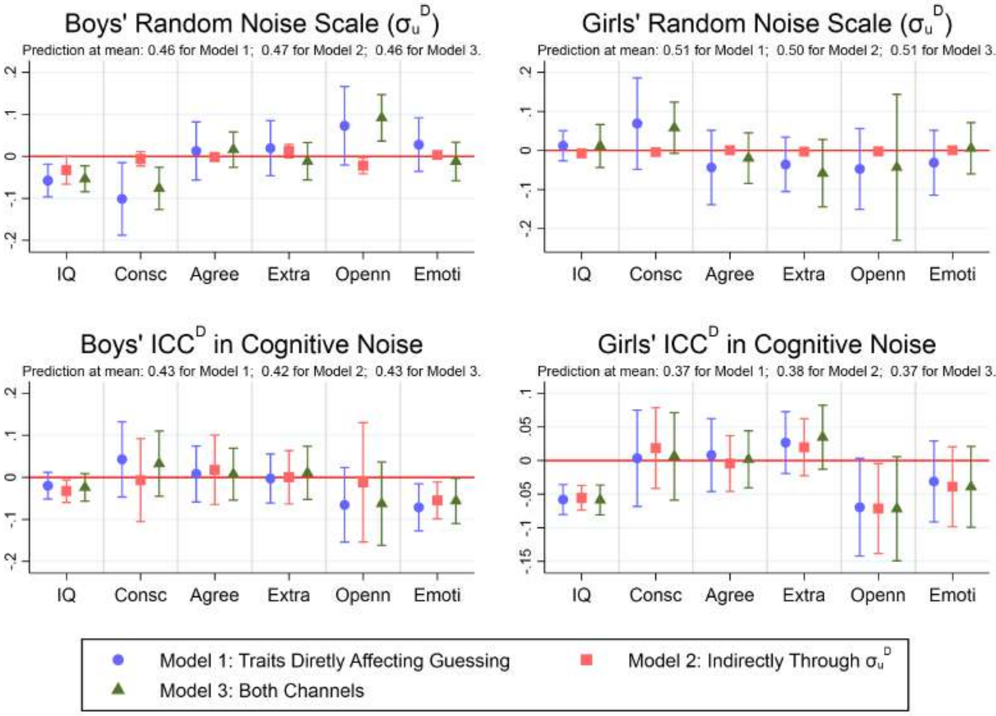


Figure I9: Random Noise Scale and ICC^D Across Guessing Models (CRRA)

Notes: Top panels show effects of IQ or personality traits on random noise scale (σ_u^D) for boys (left) and girls (right). Bottom panels show ICC^D for boys (left) and girls (right). Blue circles represent Model 1 (traits affecting guessing), red squares represent Model 2 (σ_u^D affecting guessing), green triangles represent Model 3 (both σ_u^D and traits affect guessing). Predictions at mean for σ_u^D: boys = 0.46, 0.47, 0.46, girls = 0.51, 0.50, 0.51. For ICC^D: boys = 0.43, 0.42, 0.43, girls = 0.37, 0.38, 0.37.

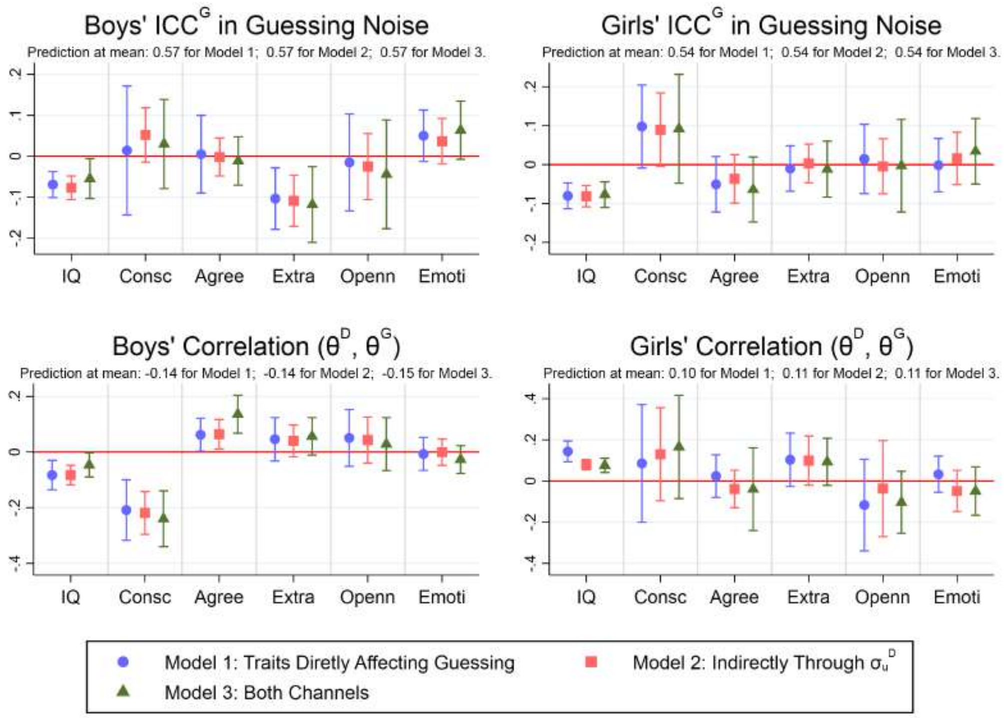


Figure I10: ICC^G and Correlation Between Random Effects Across Guessing Models (CRRA)

Notes: Top panels show ICC^G for boys (left) and girls (right). Bottom panels show correlation between deliberation and guessing random effects $\rho(\theta^D, \theta^G)$ for boys (left) and girls (right). Blue circles represent Model 1 (traits affecting guessing), red squares represent Model 2 (σ_u^D affecting guessing), green triangles represent Model 3 (both σ_u^D and traits affect guessing). Predictions at mean: ICC^G boys = 0.57, 0.57, 0.57 for Models 1–3, ICC^G girls = 0.54, 0.54, 0.54. Correlation boys = −0.14, −0.14, −0.15, girls = 0.10, 0.11, 0.11.

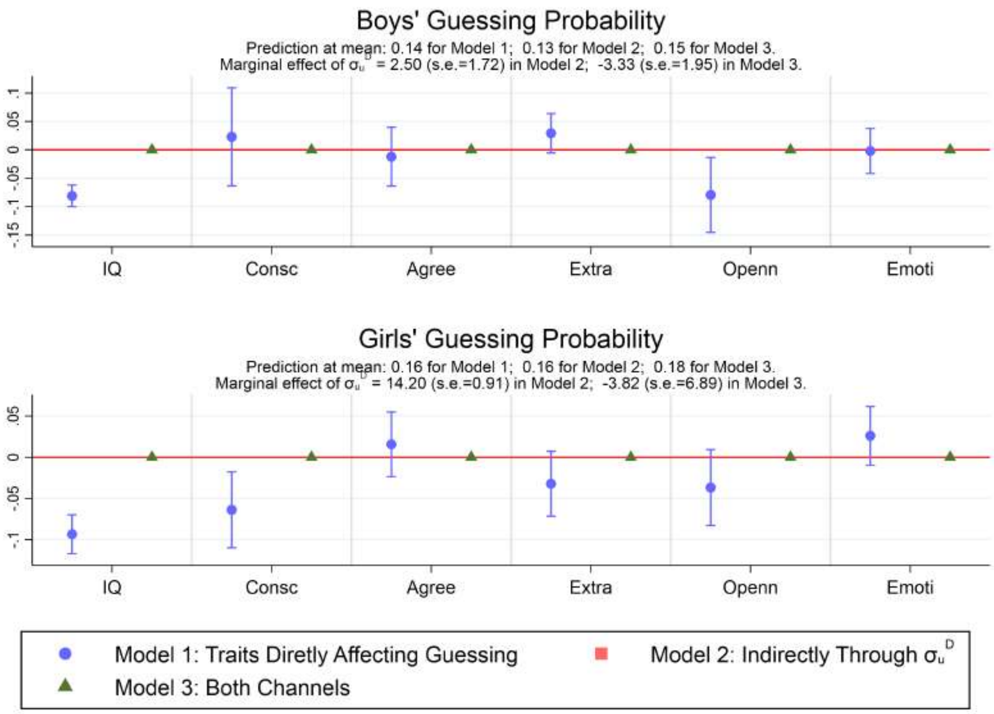


Figure I11: Guessing Probability Across Models (CRRA)

Notes: Effects of IQ or personality traits on guessing probability for boys (top) and girls (bottom). Blue circles represent Model 1 (traits affecting guessing), red squares represent Model 2 (σ_u^D affecting guessing), green triangles represent Model 3 (both σ_u^D and traits affect guessing). For Model 2, the marginal effect of σ_u^D is 2.50 (s.e. = 1.72) for boys and 14.21 (s.e. = 0.92) for girls. For Model 3, the marginal effect is -3.33 (s.e. = 1.95) for boys and -3.82 (s.e. = 6.88) for girls. Predictions at mean: boys = 0.14 for Model 1, 0.13 for Model 2, 0.15 for Model 3, girls = 0.16 for Model 1, 0.16 for Model 2, 0.18 for Model 3.

J Appendix J: Additional Figures

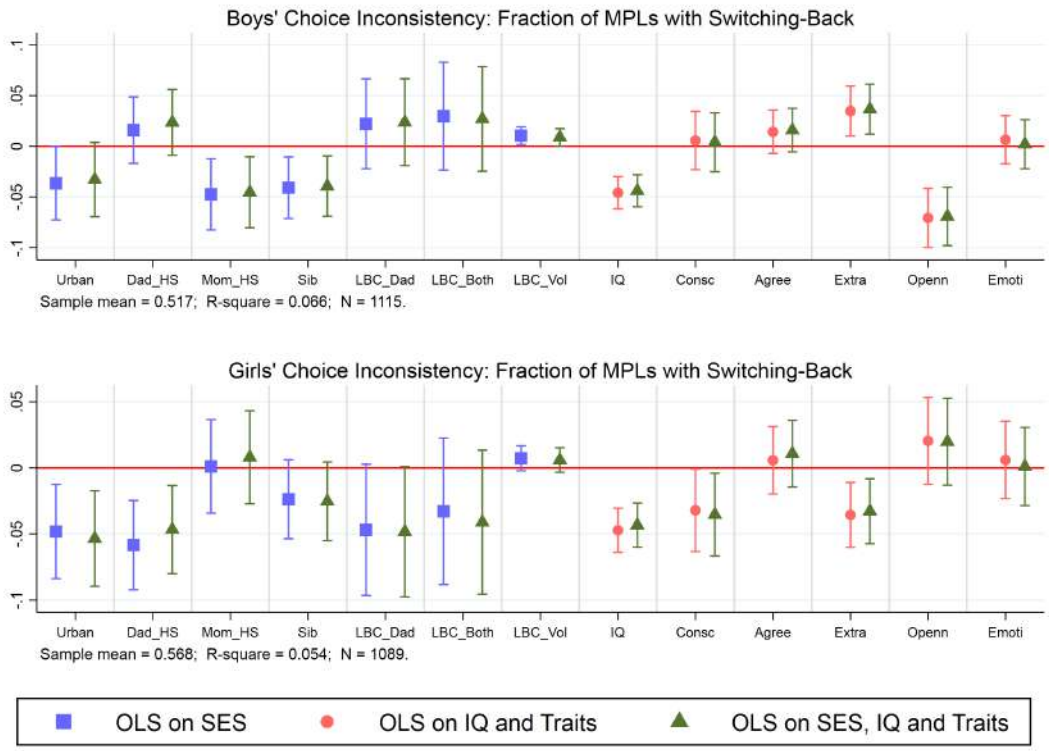


Figure J1: Reduced-Form Analysis: Choice Inconsistency (Inconsistent Behavior)

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 safer option after choosing a riskier 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 cognitive ability (IQ) and personality traits only. Green triangles represent regressions including all covariates. Error bars show 90% confidence intervals. These results correspond to Table C7 in Appendix C.

K Appendix K: 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 K1 reports effects on risk preferences and decision noise. Table K2 reports effects on ICC structure, error correlations, and guessing probability. Graphical summaries appear in the main text (Figures 2–6).

Table K1: Full Model: Effects of Cognitive and Personality Traits on Risk Preferences and Decision Noise

	Expo-Power			CRRA	Decision Noise	
Variable	r	β	$R(x,w)$ at $x=0$, $w=1$	r	σ_u^D (Expo-Power)	σ_u^D (CRRA)
Panel A: Boys						
Constant	0.770***	0.091***	0.862***	0.970***	0.556***	0.461***
	(0.033)	(0.009)	(0.028)	(0.012)	(0.041)	(0.026)
IQ	0.028	0.010	0.037	0.019	−0.103*	−0.058**
	(0.029)	(0.006)	(0.033)	(0.012)	(0.056)	(0.024)
Conscientiousness	0.104**	0.016*	0.120**	0.060***	−0.201***	−0.101*
	(0.049)	(0.009)	(0.047)	(0.022)	(0.072)	(0.053)
Agreeableness	−0.050*	0.003	−0.047	−0.003	0.065	0.013
	(0.027)	(0.009)	(0.030)	(0.020)	(0.055)	(0.042)
Extraversion	−0.021	0.004	−0.018	−0.002	0.041	0.019
	(0.027)	(0.011)	(0.029)	(0.020)	(0.053)	(0.040)
Openness	0.062	−0.043**	0.019	−0.061**	0.009	0.073
	(0.043)	(0.018)	(0.043)	(0.026)	(0.080)	(0.057)
Emotional Stability	−0.027	−0.004	−0.031	−0.015	0.056	0.028
	(0.031)	(0.010)	(0.038)	(0.020)	(0.069)	(0.039)
Panel B: Girls						
Constant	0.576***	0.093***	0.669***	0.970***	0.888***	0.509***
	(0.025)	(0.005)	(0.028)	(0.011)	(0.071)	(0.026)
IQ	−0.007	−0.005	−0.012	−0.012	0.033	0.012
	(0.024)	(0.006)	(0.029)	(0.010)	(0.071)	(0.023)
Conscientiousness	−0.049	−0.012	−0.061	−0.024	0.194	0.069
	(0.059)	(0.008)	(0.064)	(0.031)	(0.144)	(0.071)
Agreeableness	0.041	0.008	0.049	0.020	−0.129	−0.044
	(0.039)	(0.007)	(0.044)	(0.023)	(0.135)	(0.058)
Extraversion	0.090*	0.017*	0.107*	0.023	−0.252	−0.036
	(0.052)	(0.009)	(0.061)	(0.017)	(0.157)	(0.042)
Openness	−0.021	0.006	−0.015	0.016	−0.063	−0.047
	(0.056)	(0.008)	(0.061)	(0.026)	(0.148)	(0.063)
Emotional Stability	−0.025	−0.007	−0.032	0.007	0.085	−0.031
	(0.059)	(0.010)	(0.068)	(0.021)	(0.168)	(0.050)

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. For Expo-Power, $R(x,w) = r + \beta(x+w)^{1-r}$ evaluated at $x = 0$ and $w = 1$. 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 2, 3, and 4. $N = 1{,}115$ for boys and $N = 1{,}089$ for girls.

Table K2: Full Model: Effects of Cognitive and Personality Traits on ICC, Correlation, and Guessing

Variable	Expo-Power ICC^D	Expo-Power ICC^G	CRRA ICC^D	CRRA ICC^G	Both Specifications $\rho(\theta^D, \theta^G)$	Both Specifications Pr(Guess)
Panel A: Boys						
Constant	0.426***	0.580***	0.188***	0.570***	−0.151	0.127***
	(0.035)	(0.027)	(0.012)	(0.029)	(0.034)	(0.014)
IQ	−0.012	−0.078	−0.012	−0.035	−0.056	−0.070
	(0.021)	(0.022)	(0.006)	(0.022)	(0.019)	(0.009)
Conscientiousness	0.058	−0.010	0.007	−0.077	−0.234	0.038
	(0.051)	(0.064)	(0.015)	(0.050)	(0.038)	(0.031)
Agreeableness	0.010	0.005	0.008	−0.026	0.114***	−0.014
	(0.047)	(0.035)	(0.013)	(0.026)	(0.021)	(0.018)
Extraversion	−0.010	−0.079	0.006	−0.068	0.022	0.025
	(0.043)	(0.047)	(0.012)	(0.036)	(0.031)	(0.018)
Openness	−0.055	−0.065	−0.008	−0.050	0.098**	−0.058
	(0.085)	(0.058)	(0.020)	(0.049)	(0.045)	(0.024)
Emotional Stability	−0.079	0.059*	−0.001	0.041	−0.029	−0.010
	(0.039)	(0.033)	(0.010)	(0.028)	(0.029)	(0.018)
Panel B: Girls						
Constant	0.380***	0.530***	0.155***	0.524***	0.115	0.160***
	(0.029)	(0.026)	(0.010)	(0.027)	(0.074)	(0.017)
IQ	0.013	−0.034	−0.001	−0.024	0.080***	−0.100
	(0.018)	(0.020)	(0.006)	(0.021)	(0.024)	(0.012)
Conscientiousness	−0.051	0.020	0.009	0.022	0.174	−0.062
	(0.055)	(0.044)	(0.016)	(0.047)	(0.216)	(0.044)
Agreeableness	0.061	0.018	0.016	0.013	−0.060	0.021
	(0.032)	(0.028)	(0.010)	(0.029)	(0.119)	(0.036)
Extraversion	−0.044	0.008	0.007	0.016	0.081	−0.030
	(0.028)	(0.025)	(0.009)	(0.026)	(0.051)	(0.021)
Openness	−0.041	−0.021	−0.005	−0.040	−0.101	−0.025
	(0.046)	(0.033)	(0.014)	(0.034)	(0.131)	(0.029)
Emotional Stability	−0.014	0.012	0.006	0.013	−0.025	0.002
	(0.032)	(0.028)	(0.010)	(0.029)	(0.160)	(0.022)

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 decision precision. ICC^G measures the intra-class correlation in the guessing equation, capturing persistent individual differences in attentional engagement. $\rho(\theta^D, \theta^G)$ is the correlation between deliberation and guessing random effects. Pr(Guess) is the probability of guessing. The correlation and guessing probability parameters are identical across CRRA and Expo-Power specifications. Graphical representations of these coefficients appear in Figures 4, 5, and 6. N = 1,115 for boys and N = 1,089 for girls.

L Appendix L: Socioeconomic Status vs Cognitive and Personality Traits

This appendix examines whether socioeconomic status (SES) variables and cognitive/personality traits capture independent or overlapping sources of heterogeneity in risk preferences and decision quality. We compare three specifications under both CRRA and Expo-Power: Model 1 includes only SES variables (urbanicity, parental education, sibling structure, 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 L9–L14 present results for CRRA, while Figures L1–L8 present results for Expo-Power.

Expo-Power Specification

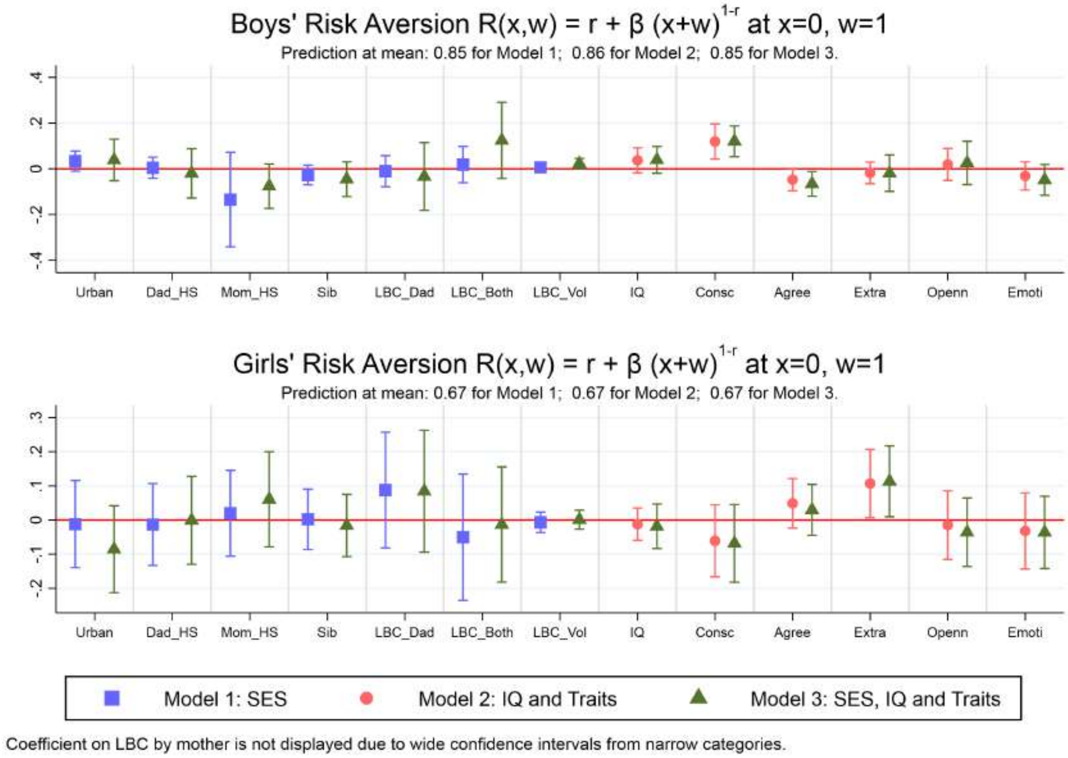


Figure L1: SES vs Traits: Risk Aversion under Expo-Power

Notes: Overall risk aversion $R(x, w) = r + \beta(x + w)^{1-r}$ evaluated at $x = 0$ and $w = 1$ across three covariate specifications under Expo-Power. Top panel: boys, bottom panel: girls. Blue squares: Model 1 (SES only), red circles: Model 2 (IQ and traits only), green triangles: Model 3 (SES, IQ, and traits). Predictions at mean: boys show $R(x, w) = 0.85$ (Model 1), 0.86 (Model 2), 0.85 (Model 3); girls show $R(x, w) = 0.67$ across all models. Error bars show 90% confidence intervals. Coefficient on LBC by mother not displayed due to wide confidence intervals from narrow categories. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

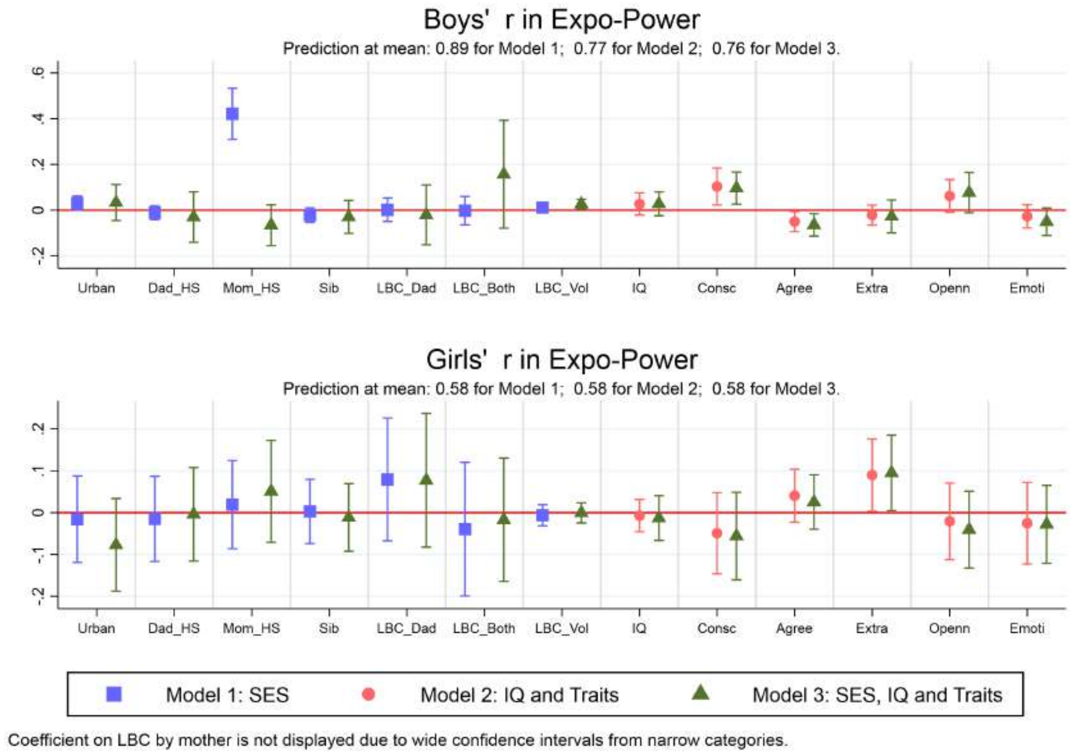


Figure L2: SES vs Traits: Baseline Curvature r under Expo-Power

Notes: Baseline curvature parameter r across covariate specifications under Expo-Power. Predictions at mean: boys show $r = 0.89$ (Model 1), 0.77 (Model 2), 0.76 (Model 3); girls show $r = 0.58$ across all models. For boys, the shift from Model 1 to Models 2 and 3 reflects offsetting movements with β (Figure L3) that leave overall risk aversion $R(x, w)$ unchanged. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

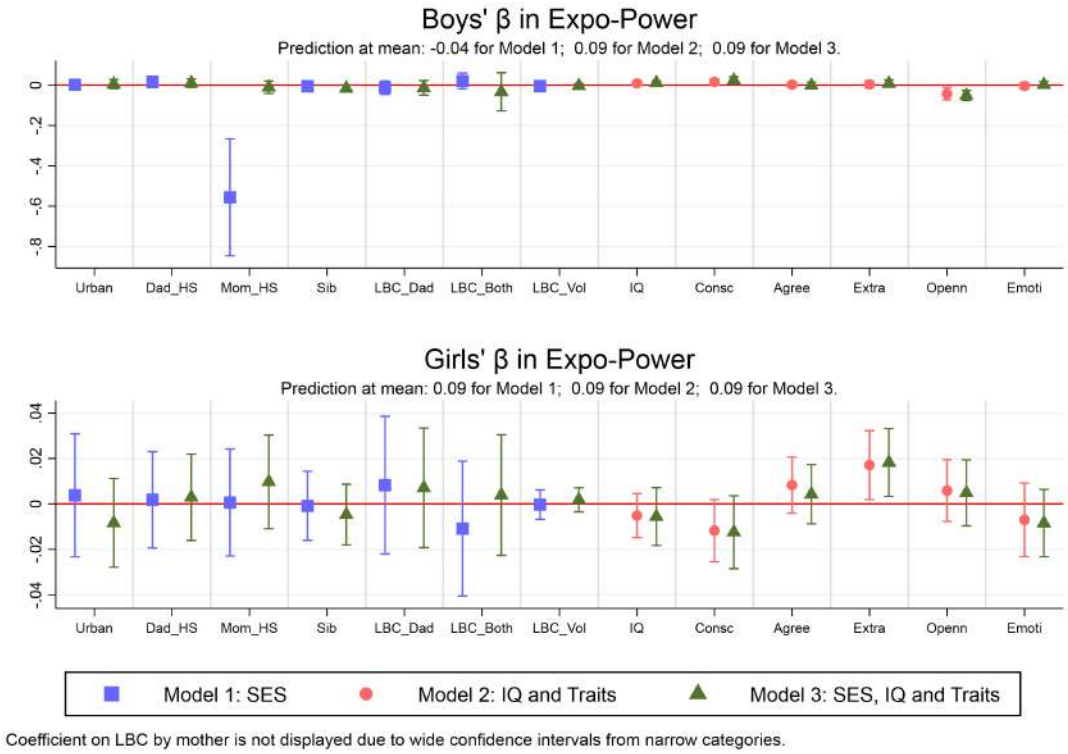


Figure L3: SES vs Traits: Wealth Sensitivity β under Expo-Power

Notes: Wealth sensitivity parameter β across covariate specifications under Expo-Power. Predictions at mean: boys show $\beta = -0.04$ (Model 1), 0.09 (Models 2 and 3); girls show $\beta = 0.09$ across all models. For boys, SES variables alone yield negative β, while adding IQ and traits shifts β positive with corresponding reduction in r (Figure L2), such that overall risk aversion remains constant. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

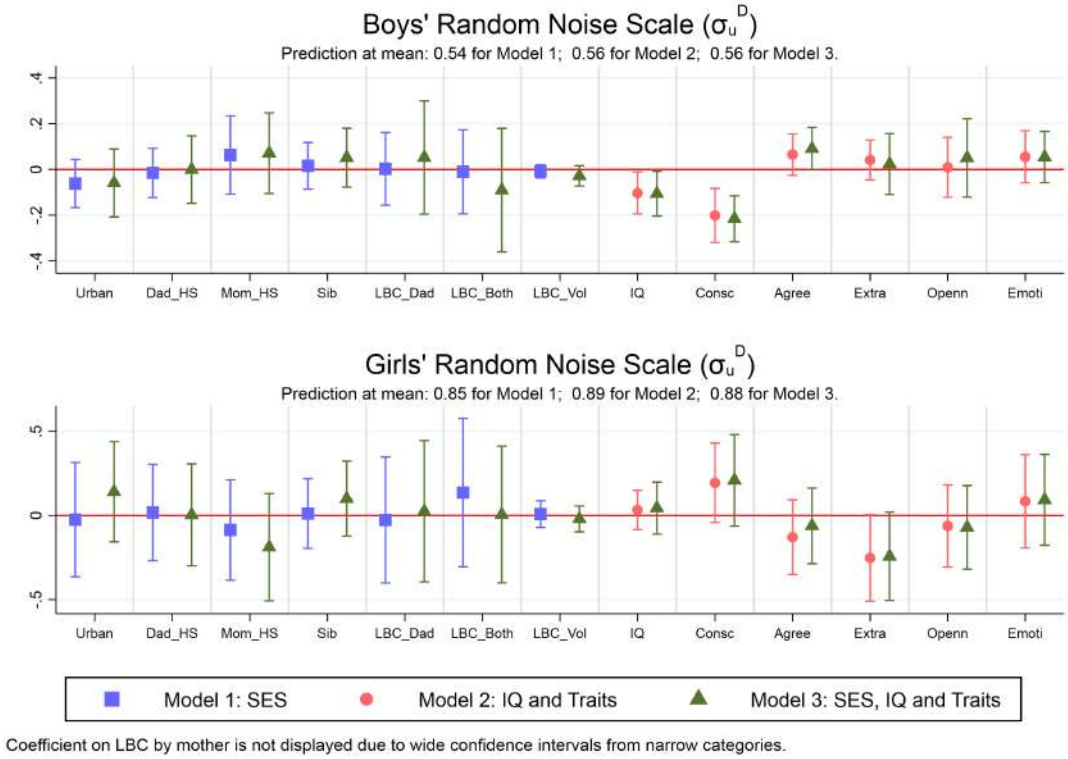


Figure L4: SES vs Traits: Random Noise Scale under Expo-Power

Notes: Transitory noise σ_u^D across covariate specifications under Expo-Power. Predictions at mean: boys show $\sigma_u^D = 0.54$ (Model 1), 0.56 (Models 2 and 3); girls show $\sigma_u^D = 0.85$ (Model 1), 0.89 (Model 2), 0.88 (Model 3). Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

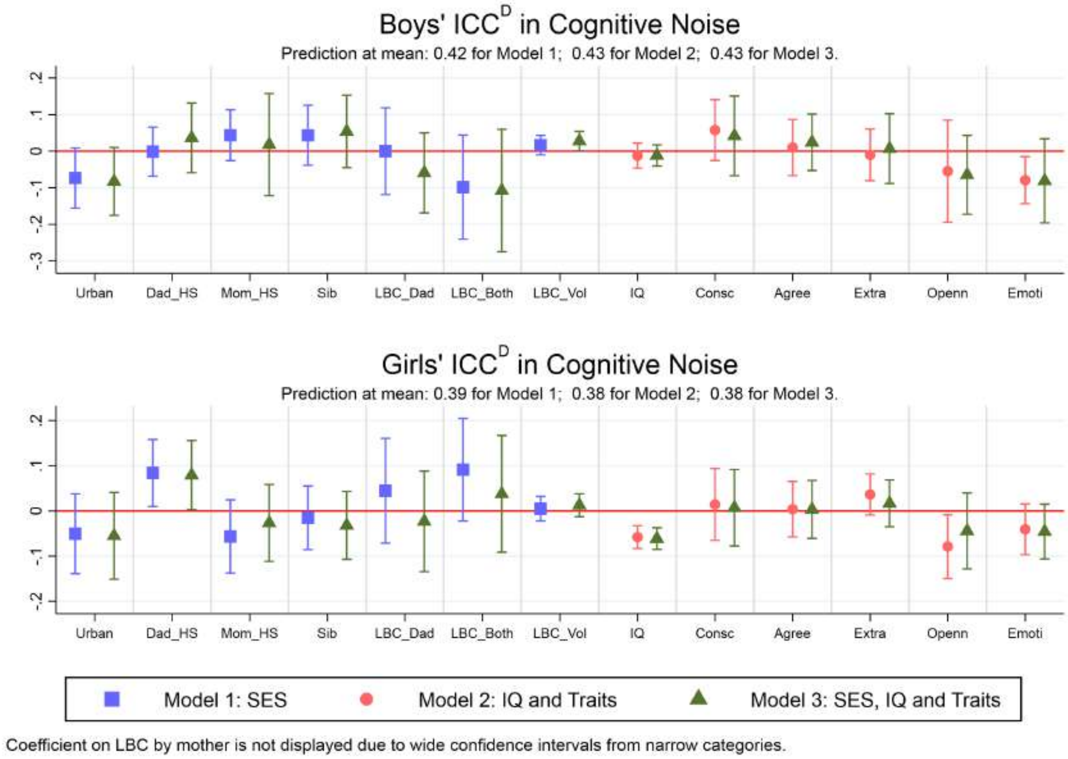


Figure L5: SES vs Traits: ICC^{D} under Expo-Power

Notes: ICC^{D} across covariate specifications under Expo-Power. Predictions at mean: boys show ICC^{D} = 0.42 (Model 1), 0.43 (Models 2 and 3); girls show ICC^{D} = 0.39 (Model 1), 0.38 (Models 2 and 3). Error bars show 90% confidence intervals. N = 1,115 boys, N = 1,089 girls.

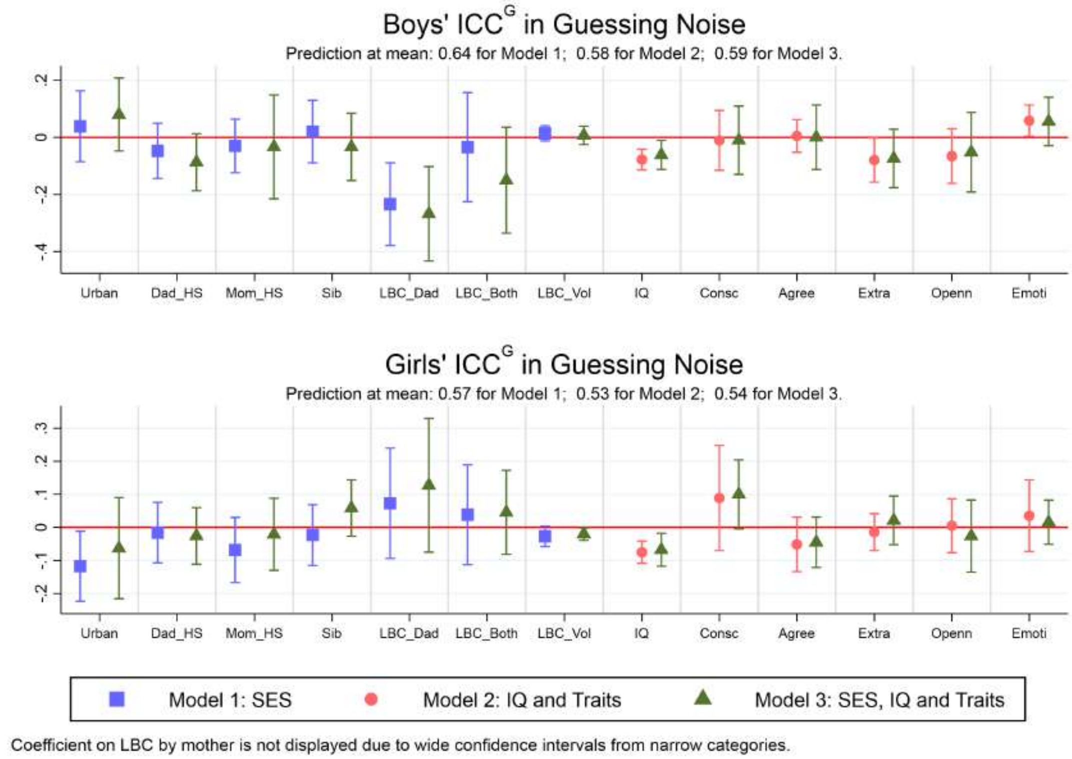


Figure L6: SES vs Traits: ICC^G under Expo-Power

Notes: ICC^G across covariate specifications under Expo-Power. Predictions at mean: boys show ICC^G = 0.64 (Model 1), 0.58 (Model 2), 0.59 (Model 3); girls show ICC^G = 0.57 (Model 1), 0.53 (Model 2), 0.54 (Model 3). As under CRRA, boys show approximately 10% reduction when adding IQ and traits. Error bars show 90% confidence intervals. N = 1,115 boys, N = 1,089 girls.

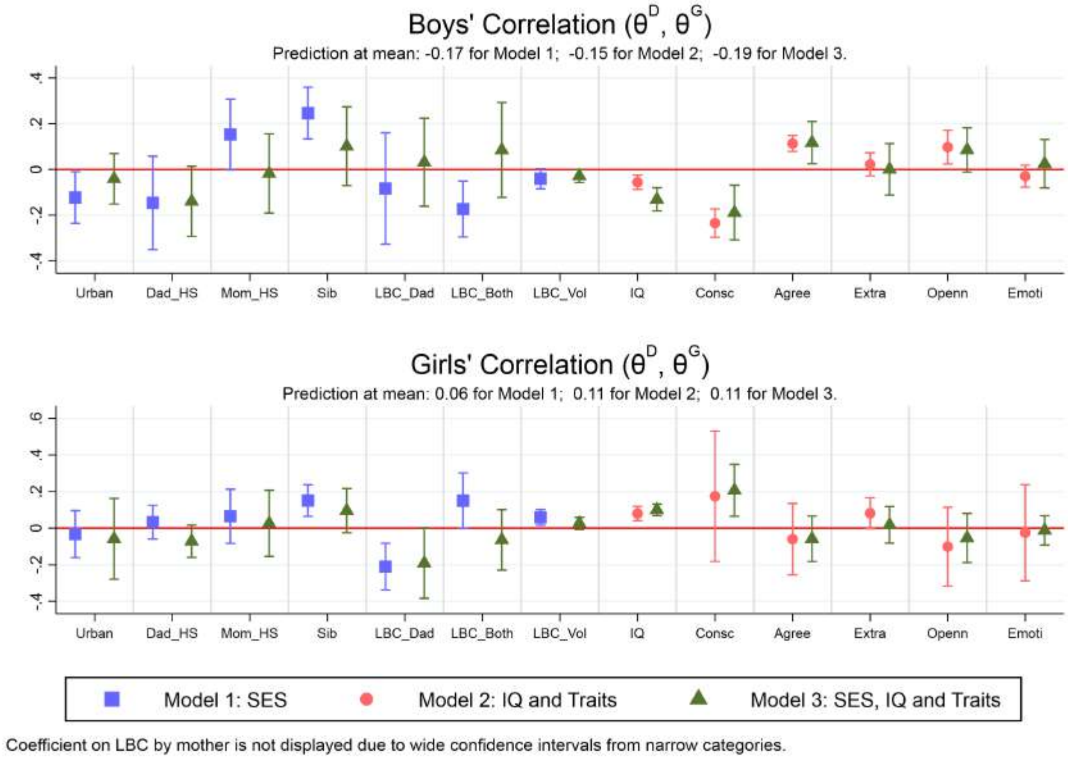


Figure L7: SES vs Traits: Correlation $\rho(\theta^D, \theta^G)$ under Expo-Power

Notes: Correlation between deliberation and guessing random effects $\rho(\theta^D, \theta^G)$ across covariate specifications under Expo-Power. Predictions at mean: boys show $\rho = -0.17$ (Model 1), -0.15 (Model 2), -0.19 (Model 3); girls show $\rho = 0.06$ (Model 1), 0.11 (Models 2 and 3). The opposite-signed correlations by gender persist across all specifications and both functional forms. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

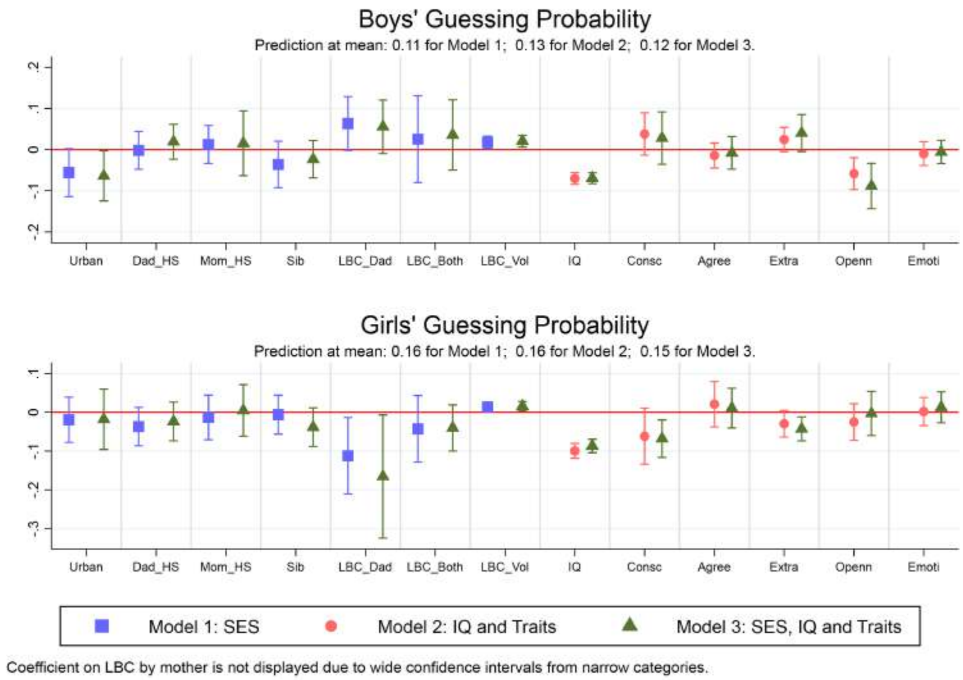


Figure L8: SES vs Traits: Guessing Probability under Expo-Power

Notes: Guessing probability across covariate specifications under Expo-Power. Predictions at mean: boys show Pr(Guess) = 0.11 (Model 1), 0.13 (Model 2), 0.12 (Model 3); girls show Pr(Guess) = 0.16 (Model 1), 0.16 (Model 2), 0.15 (Model 3). Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

CRRA Specification

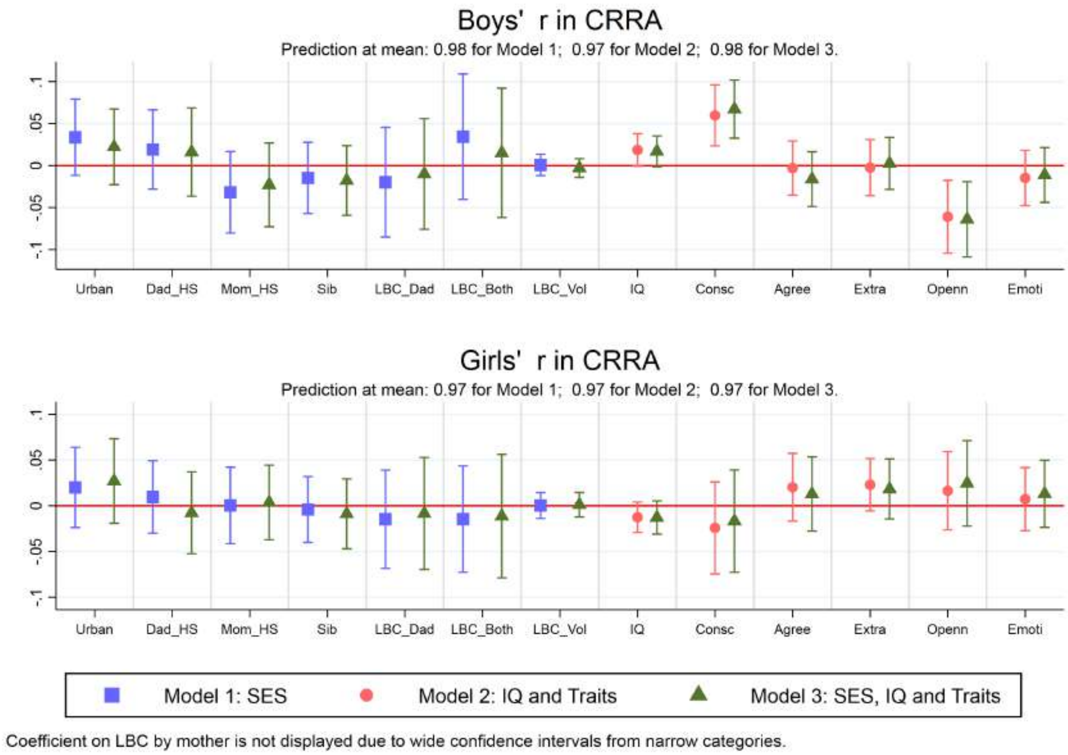


Figure L9: SES vs Traits: Risk Aversion under CRRA

Notes: Risk aversion r across three covariate specifications under CRRA. Top panel: boys, bottom panel: girls. Blue squares: Model 1 (SES only), red circles: Model 2 (IQ and traits only), green triangles: Model 3 (SES, IQ, and traits). Predictions at mean: boys show $r = 0.98$ (Model 1), 0.97 (Model 2), 0.98 (Model 3); girls show $r = 0.97$ across all models. Error bars show 90% confidence intervals. Coefficient on LBC by mother not displayed due to wide confidence intervals from narrow categories. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

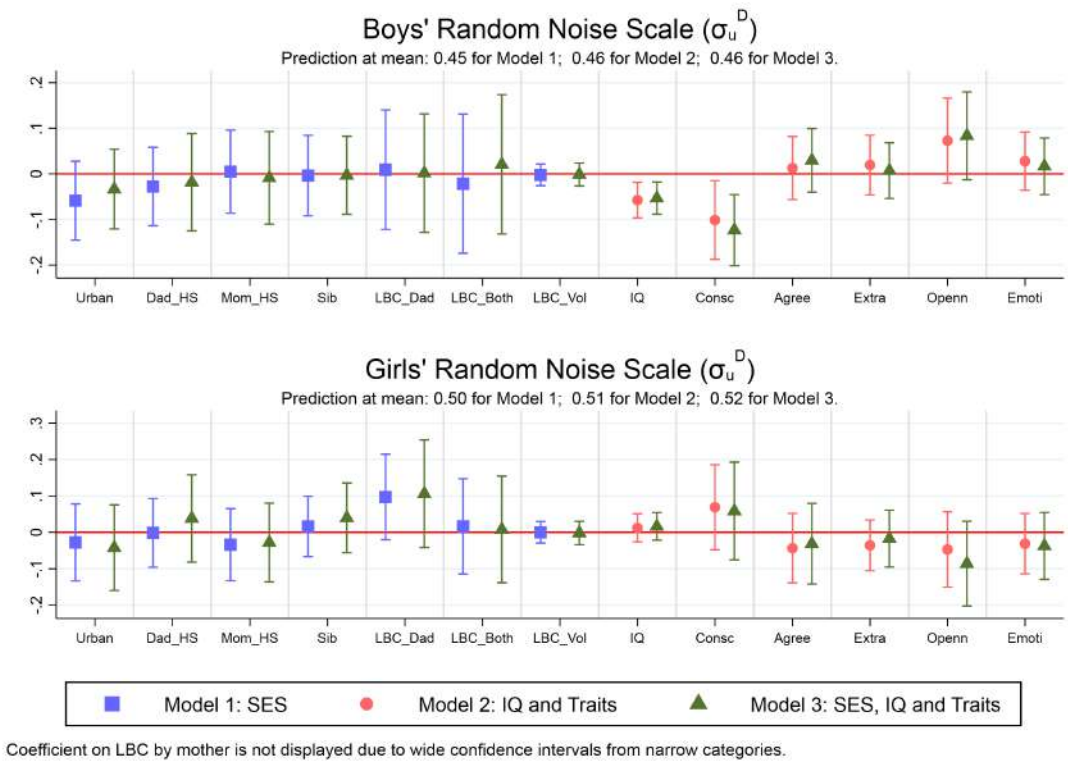


Figure L10: SES vs Traits: Random Noise Scale under CRRA

Notes: Transitory noise σ_u^D across covariate specifications under CRRA. Predictions at mean: boys show $\sigma_u^D = 0.45$ (Model 1), 0.46 (Models 2 and 3); girls show $\sigma_u^D = 0.50$ (Model 1), 0.51 (Model 2), 0.52 (Model 3). Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

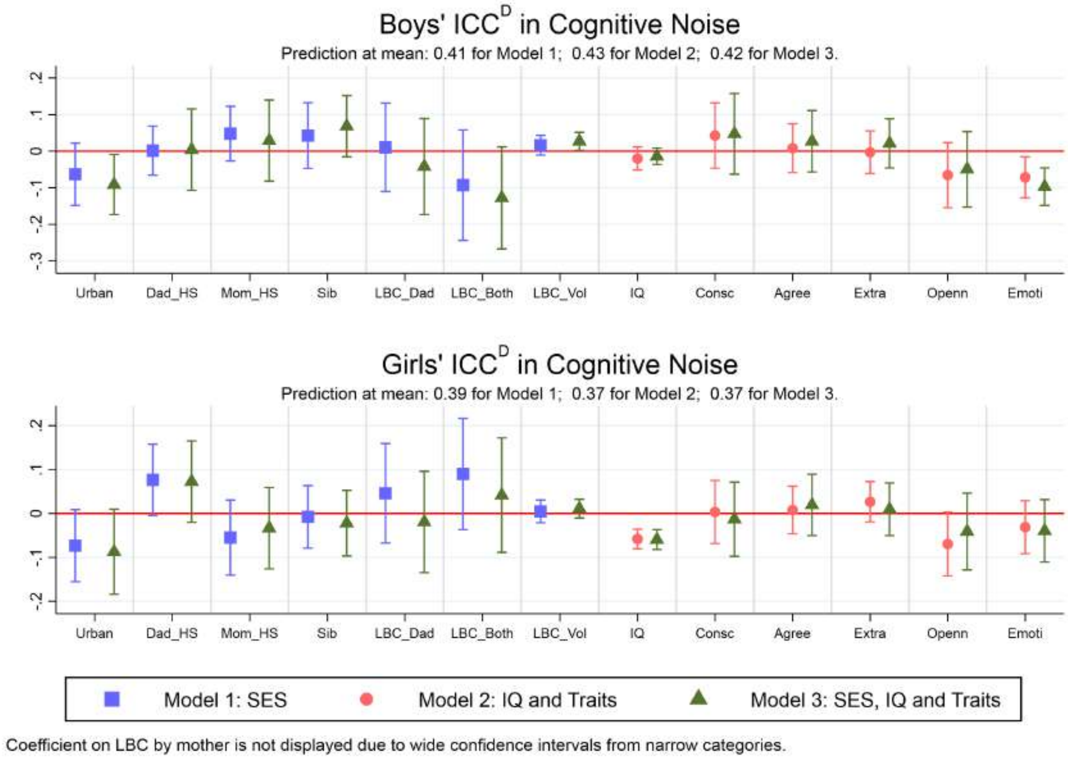


Figure L11: SES vs Traits: ICC^D under CRRA

Notes: ICC^D across covariate specifications under CRRA. Predictions at mean: boys show ICC^D = 0.41 (Model 1), 0.43 (Models 2 and 3); girls show ICC^D = 0.39 (Model 1), 0.37 (Models 2 and 3). Error bars show 90% confidence intervals. N = 1,115 boys, N = 1,089 girls.

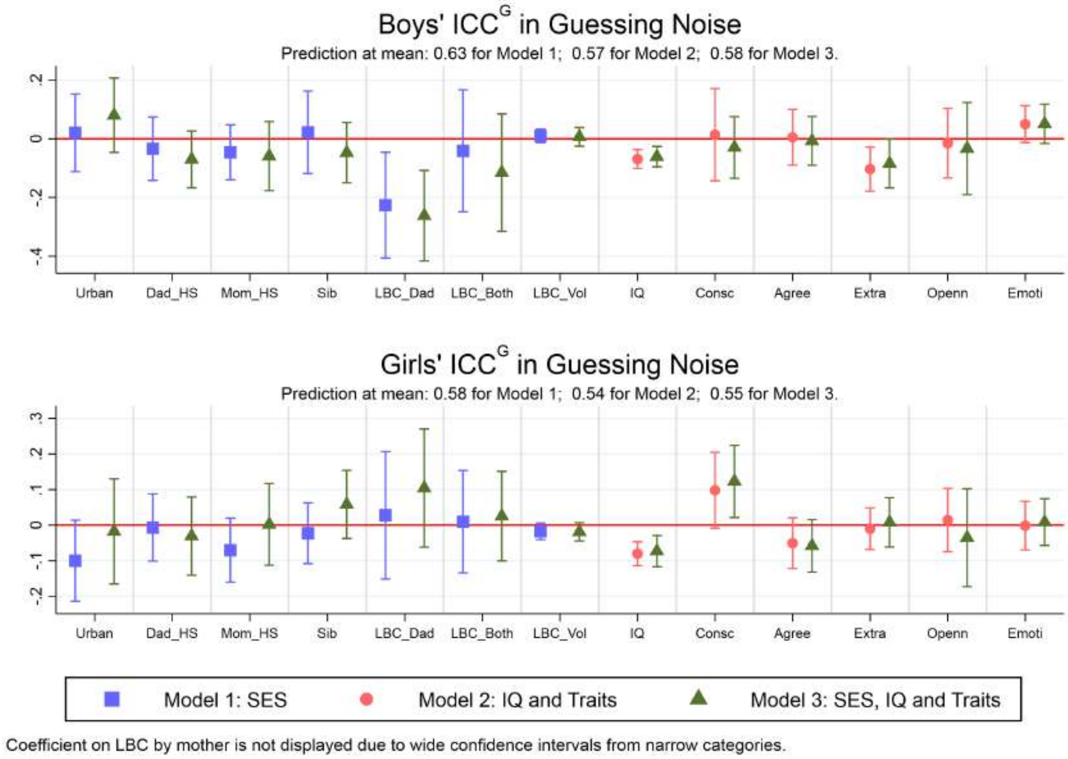


Figure L12: SES vs Traits: ICC^G under CRRA

Notes: ICC^G across covariate specifications under CRRA. Predictions at mean: boys show ICC^G = 0.63 (Model 1), 0.57 (Models 2 and 3); girls show ICC^G = 0.58 (Model 1), 0.54 (Model 2), 0.55 (Model 3). Boys show a 10% reduction when adding IQ and traits, reflecting the predictive power of personality traits—particularly Emotional Stability—for persistent guessing patterns. Error bars show 90% confidence intervals. N = 1,115 boys, N = 1,089 girls.

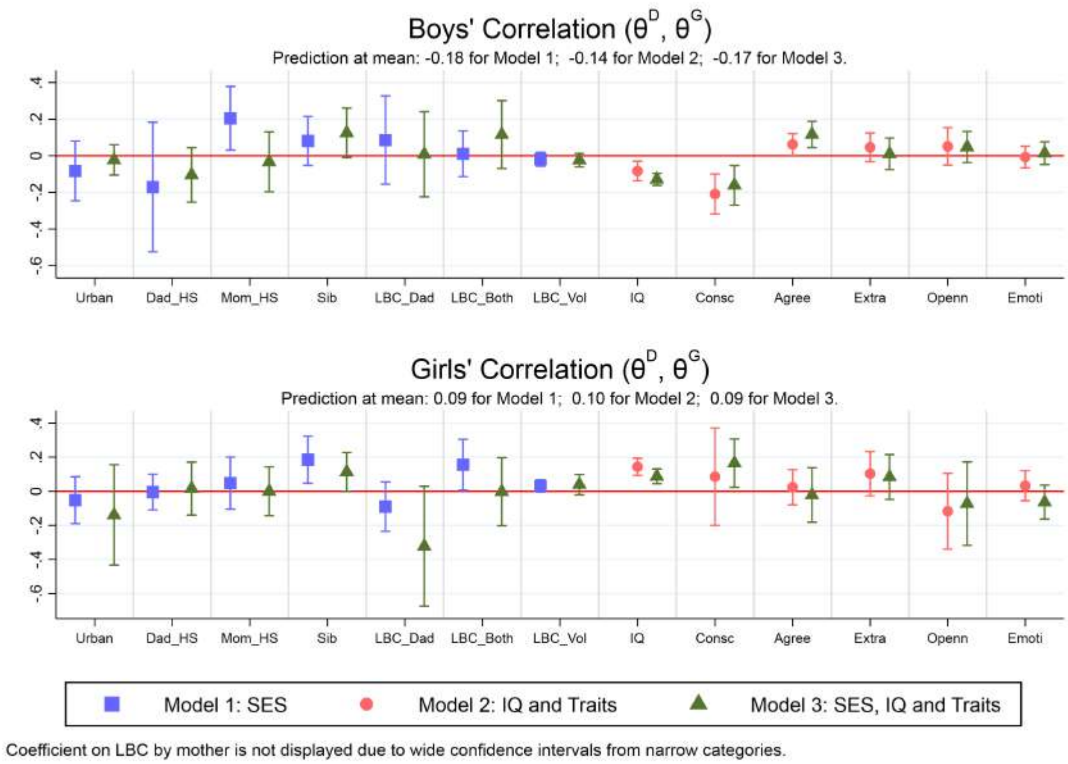


Figure L13: SES vs Traits: Correlation $\rho(\theta^D, \theta^G)$ under CRRA

Notes: Correlation between deliberation and guessing random effects $\rho(\theta^D, \theta^G)$ across covariate specifications under CRRA. Predictions at mean: boys show $\rho = -0.18$ (Model 1), -0.14 (Models 2 and 3); girls show $\rho = 0.09$ (Model 1), 0.10 (Model 2), 0.09 (Model 3). The opposite-signed correlations by gender persist across all specifications. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

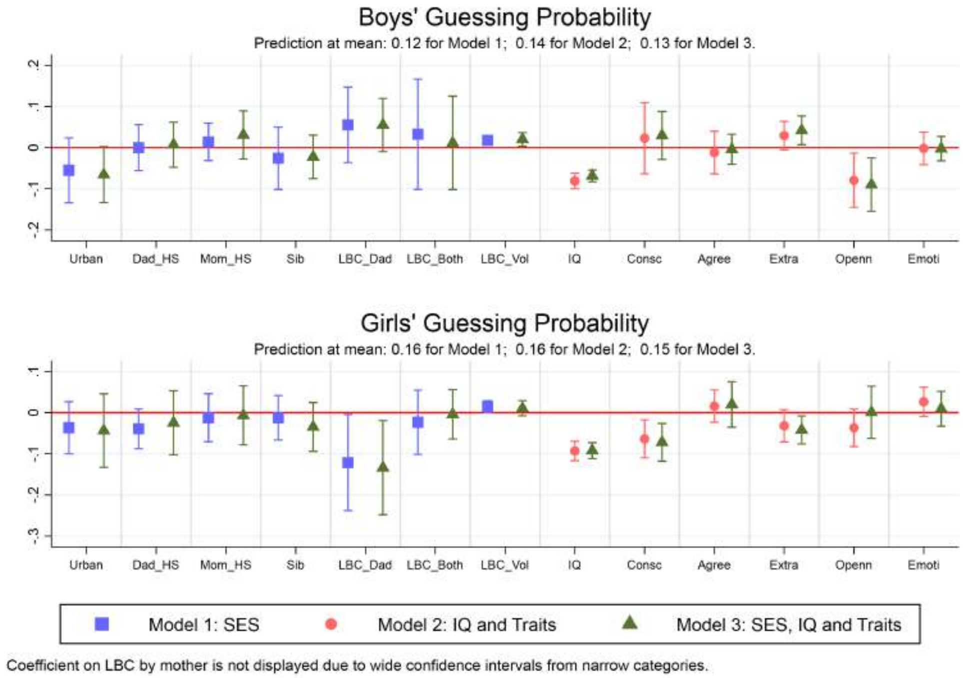


Figure L14: SES vs Traits: Guessing Probability under CRRA

Notes: Guessing probability across covariate specifications under CRRA. Predictions at mean: boys show Pr(Guess) = 0.12 (Model 1), 0.14 (Model 2), 0.13 (Model 3); girls show Pr(Guess) = 0.16 across all models. Error bars show 90% confidence intervals. $N = 1{,}115$ boys, $N = 1{,}089$ girls.

M Appendix M: Sensitivity of Estimates to the Assumed Fallback Consumption Level

This appendix examines the sensitivity of our estimates to the assumed fallback consumption level w_0, which represents the consumption level when an individual loses everything or gains nothing from the game. In our main analyses, this level is assumed to be fixed at $w_0 = 1$. Here we relax this assumption to examine whether our estimates are sensitive to this normalization.

We re-estimate both CRRA and Expo-Power models fixing w_0 at values around 1—specifically, 0.2, 0.5, 1, 2, and 5. For each assumed fallback level, total consumption becomes $x = w_0 + x_1$, where x_1 is the payoff from the experimental game. This affects how relative risk aversion $R(x, w)$ is evaluated and how the curvature parameters translate into choice behavior.

We perform this sensitivity check for both the Expo-Power (Figures M1–M6) and CRRA (Figures M7–M11) utility functions. Each figure displays coefficient estimates across these five fallback consumption levels: blue triangles ($w_0 = 0.2$), red squares ($w_0 = 0.5$), green circles ($w_0 = 1$, our baseline), orange squares ($w_0 = 2$), and purple triangles ($w_0 = 5$).

Despite shifts in the marginal effects of IQ and personality traits across alternative values of the fallback consumption level w_0, our 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.

M.1 Expo-Power Specification

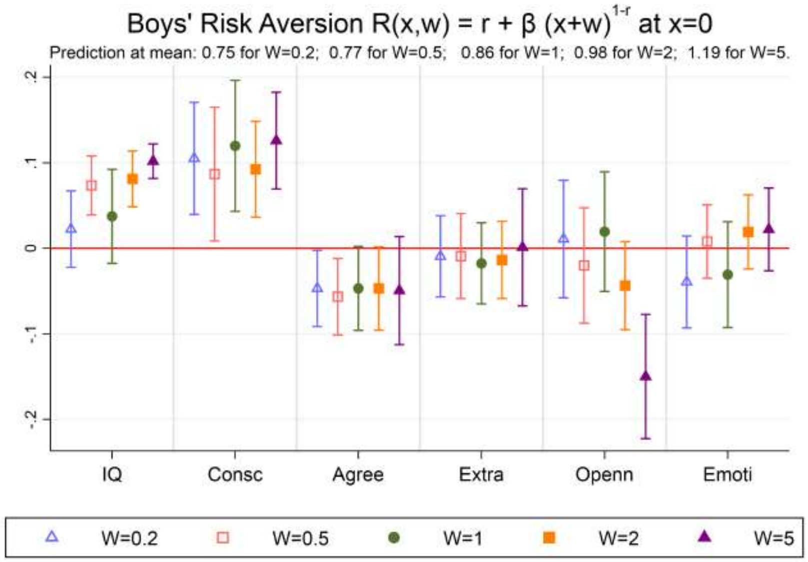


Figure M1: Risk Aversion Across Fallback Consumption Levels in Expo-Power: Boys

Notes: Effects of IQ or personality traits on risk aversion $R(x, w) = r + \beta(x + w)^{1-r}$ evaluated at $x = 0$ and $w = 1$ for boys across fallback consumption levels. Blue triangles represent $w_0 = 0.2$, red squares $w_0 = 0.5$, green circles $w_0 = 1$, orange squares $w_0 = 2$, and purple triangles $w_0 = 5$. Predictions at mean: 0.75, 0.77, 0.86, 0.98, 1.19 across fallback levels.

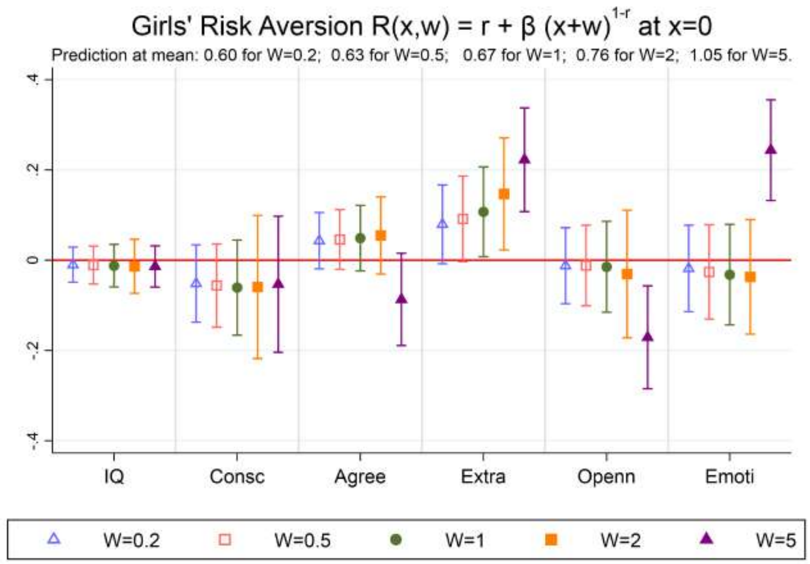


Figure M2: Risk Aversion Across Fallback Consumption Levels in Expo-Power: Girls

Notes: Effects of IQ or personality traits on risk aversion $R(x, w) = r + \beta(x + w)^{1-r}$ evaluated at $x = 0$ and $w = 1$ for girls across fallback consumption levels. Blue triangles represent $w_0 = 0.2$, red squares $w_0 = 0.5$, green circles $w_0 = 1$, orange squares $w_0 = 2$, and purple triangles $w_0 = 5$. Predictions at mean: 0.60, 0.63, 0.67, 0.76, 1.05 across fallback levels.

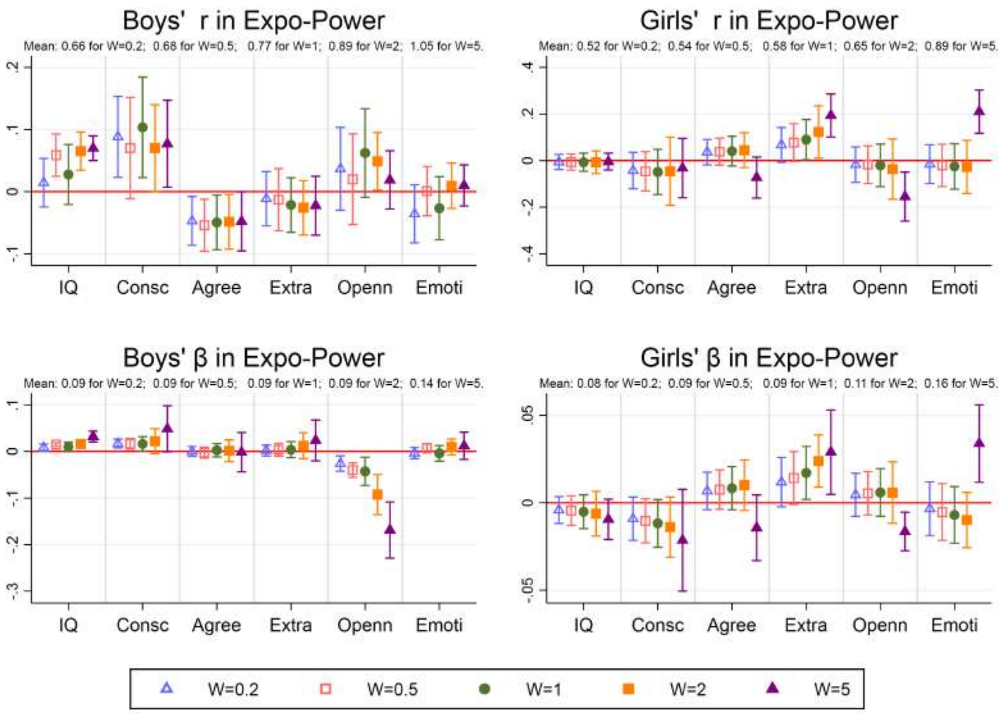


Figure M3: Expo-Power Parameters r and β Across Fallback Consumption Levels

Notes: Top panels show effects of IQ or personality traits on parameter r for boys (left) and girls (right). Bottom panels show effects on parameter β. Blue triangles represent $w_0 = 0.2$, red squares $w_0 = 0.5$, green circles $w_0 = 1$, orange squares $w_0 = 2$, and purple triangles $w_0 = 5$. Predictions at mean for r: boys = 0.66, 0.68, 0.77, 0.89, 1.05, girls = 0.52, 0.54, 0.58, 0.65, 0.89. For β: boys = 0.09, 0.09, 0.09, 0.09, 0.14, girls = 0.08, 0.09, 0.09, 0.11, 0.16.

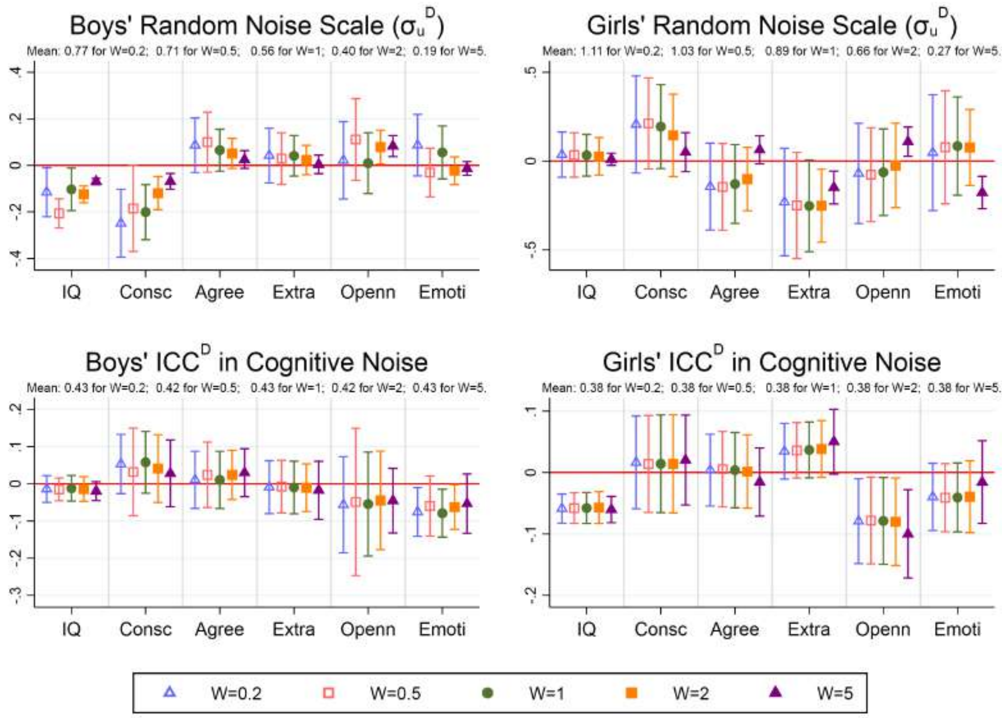


Figure M4: Transitory Noise Scale and ICC^D Across Fallback Consumption Levels in Expo-Power

Notes: Top panels show effects of IQ or personality traits on transitory noise scale (σ_u^D) for boys (left) and girls (right). Bottom panels show ICC^D. Blue triangles represent $w_0 = 0.2$, red squares $w_0 = 0.5$, green circles $w_0 = 1$, orange squares $w_0 = 2$, and purple triangles $w_0 = 5$. Predictions at mean for σ_u^D: boys = 0.77, 0.71, 0.56, 0.40, 0.19, girls = 1.11, 1.03, 0.89, 0.66, 0.27. For ICC^D: boys = 0.43, 0.42, 0.43, 0.42, 0.43, girls = 0.38, 0.38, 0.38, 0.38, 0.38.

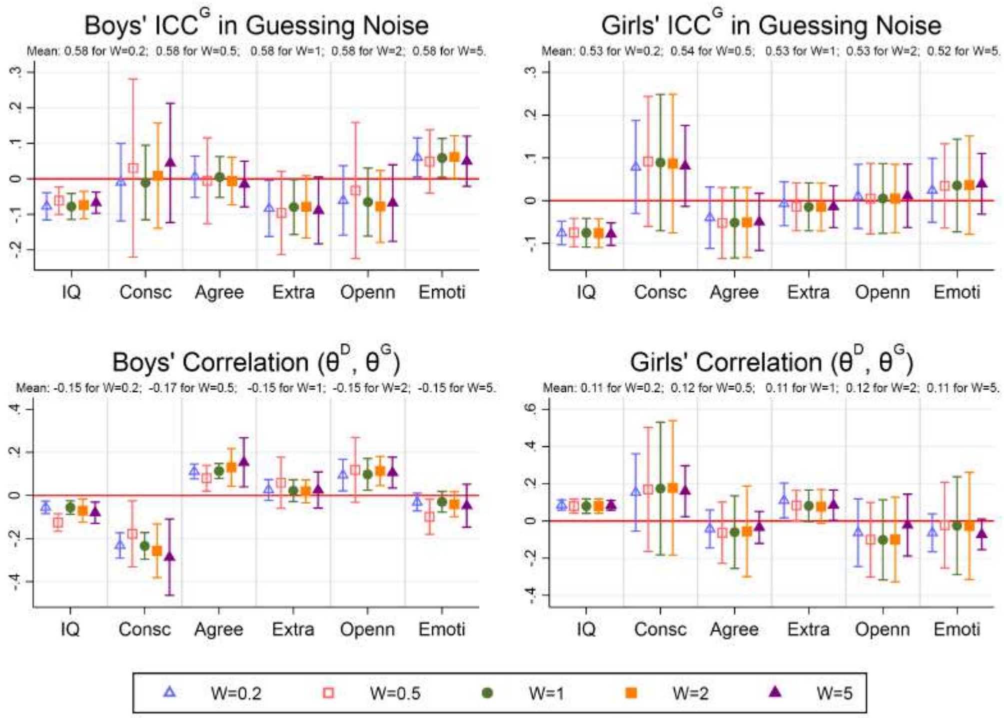


Figure M5: ICC^G and Correlation Between Persistent Noise Components Across Fallback Consumption Levels in Expo-Power

Notes: Top panels show ICC^G for boys (left) and girls (right). Bottom panels show correlation between deliberation and guessing persistent noise components $\rho(\theta^D, \theta^G)$. Blue triangles represent $w_0 = 0.2$, red squares $w_0 = 0.5$, green circles $w_0 = 1$, orange squares $w_0 = 2$, and purple triangles $w_0 = 5$. Predictions at mean: ICC^G boys = 0.58, 0.58, 0.58, 0.58, 0.58, ICC^G girls = 0.53, 0.54, 0.53, 0.53, 0.52. Correlation boys = -0.15, -0.17, -0.15, -0.15, -0.15, girls = 0.11, 0.12, 0.11, 0.12, 0.11.

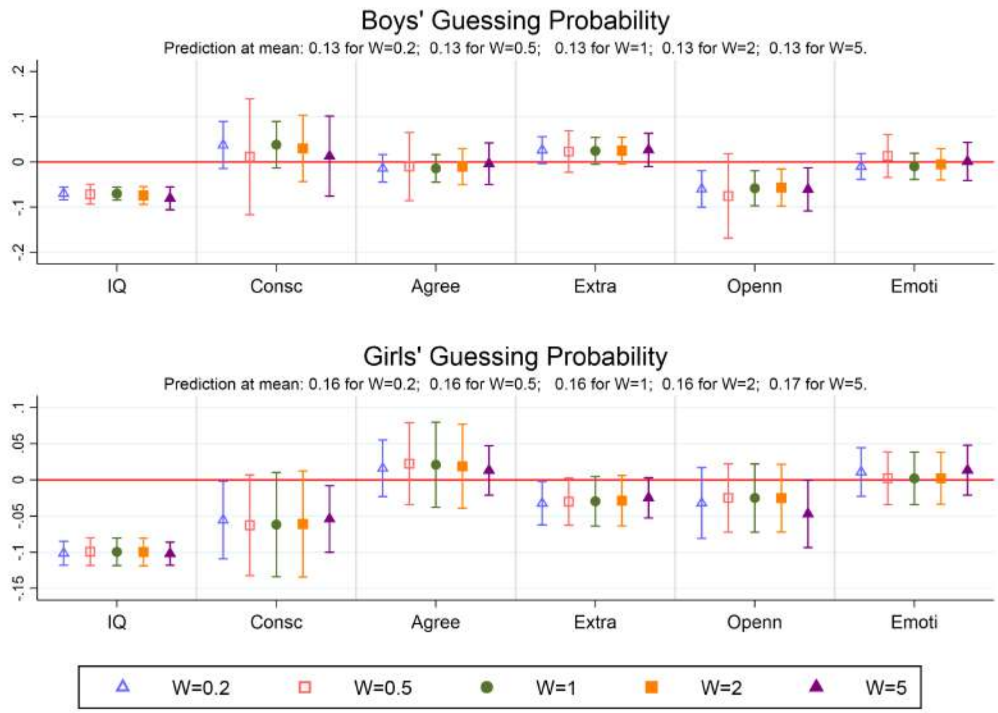


Figure M6: Guessing Probability Across Fallback Consumption Levels in Expo-Power

Notes: Effects of IQ or personality traits on guessing probability for boys (top) and girls (bottom) under the Expo-Power specification. Blue triangles represent $w_0 = 0.2$, red squares $w_0 = 0.5$, green circles $w_0 = 1$, orange squares $w_0 = 2$, and purple triangles $w_0 = 5$. Predictions at mean: boys $= 0.13, 0.13, 0.13, 0.13, 0.13$, girls $= 0.16, 0.16, 0.16, 0.16, 0.17$. The stability across fallback consumption levels confirms that our main findings are robust to this normalization choice.

M.2 CRRA Specification

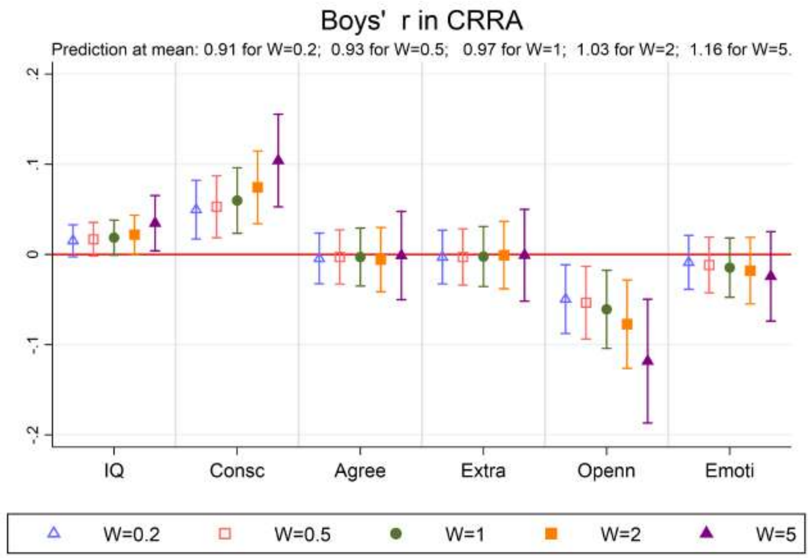


Figure M7: Risk Aversion Parameter r Across Fallback Consumption Levels in CRRA: Boys

Notes: Effects of IQ or personality traits on CRRA parameter r for boys across fallback consumption levels. Blue triangles represent $w_0 = 0.2$, red squares $w_0 = 0.5$, green circles $w_0 = 1$, orange squares $w_0 = 2$, and purple triangles $w_0 = 5$. Predictions at mean: 0.91, 0.93, 0.97, 1.03, 1.16 across fallback levels.

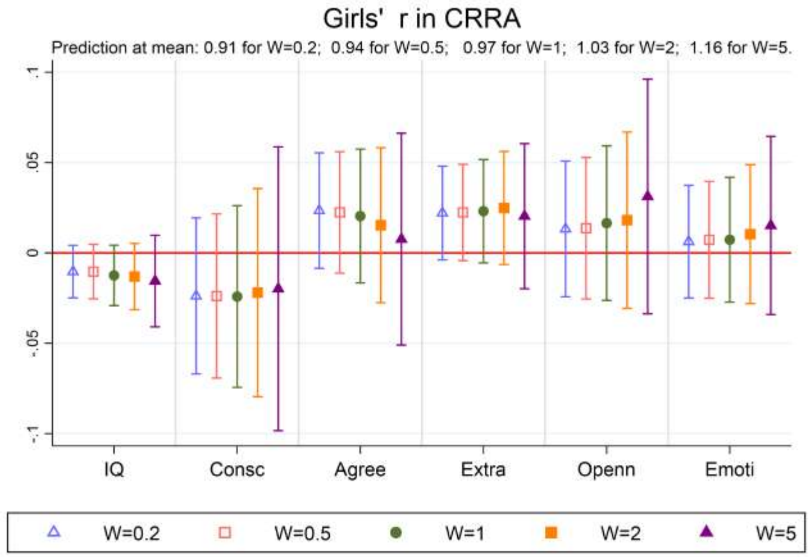


Figure M8: Risk Aversion Parameter r Across Fallback Consumption Levels in CRRA: Girls

Notes: Effects of IQ or personality traits on CRRA parameter r for girls across fallback consumption levels. Blue triangles represent $w_0 = 0.2$, red squares $w_0 = 0.5$, green circles $w_0 = 1$, orange squares $w_0 = 2$, and purple triangles $w_0 = 5$. Predictions at mean: 0.91, 0.94, 0.97, 1.03, 1.16 across fallback levels.

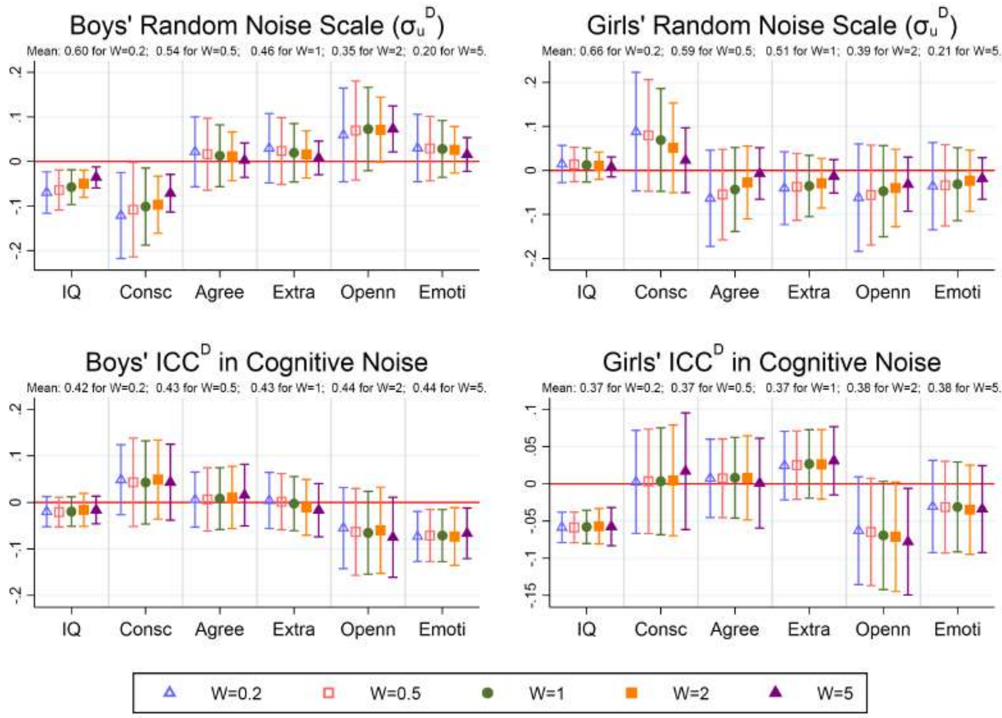


Figure M9: Transitory Noise Scale and ICC^D Across Fallback Consumption Levels in CRRA

Notes: Top panels show effects of IQ or personality traits on transitory noise scale (σ_u^D) for boys (left) and girls (right). Bottom panels show ICC^D. Blue triangles represent $w_0 = 0.2$, red squares $w_0 = 0.5$, green circles $w_0 = 1$, orange squares $w_0 = 2$, and purple triangles $w_0 = 5$. Predictions at mean for σ_u^D: boys = 0.60, 0.54, 0.46, 0.35, 0.20, girls = 0.66, 0.59, 0.51, 0.39, 0.21. For ICC^D: boys = 0.42, 0.43, 0.43, 0.44, 0.44, girls = 0.37, 0.37, 0.37, 0.38, 0.38.

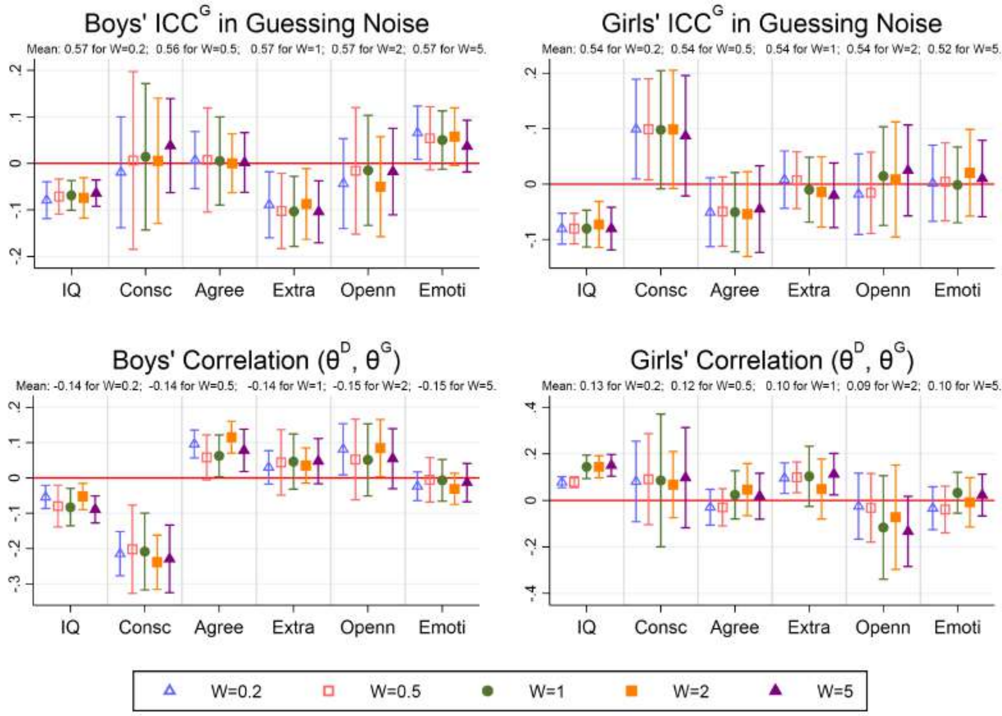


Figure M10: ICC^G and Correlation Between Persistent Noise Components Across Fallback Consumption Levels in CRRA

Notes: Top panels show ICC^G for boys (left) and girls (right). Bottom panels show correlation between deliberation and guessing persistent noise components $\rho(\theta^D, \theta^G)$. Blue triangles represent $w_0 = 0.2$, red squares $w_0 = 0.5$, green circles $w_0 = 1$, orange squares $w_0 = 2$, and purple triangles $w_0 = 5$. ICC and correlation parameters remain stable across fallback consumption levels.

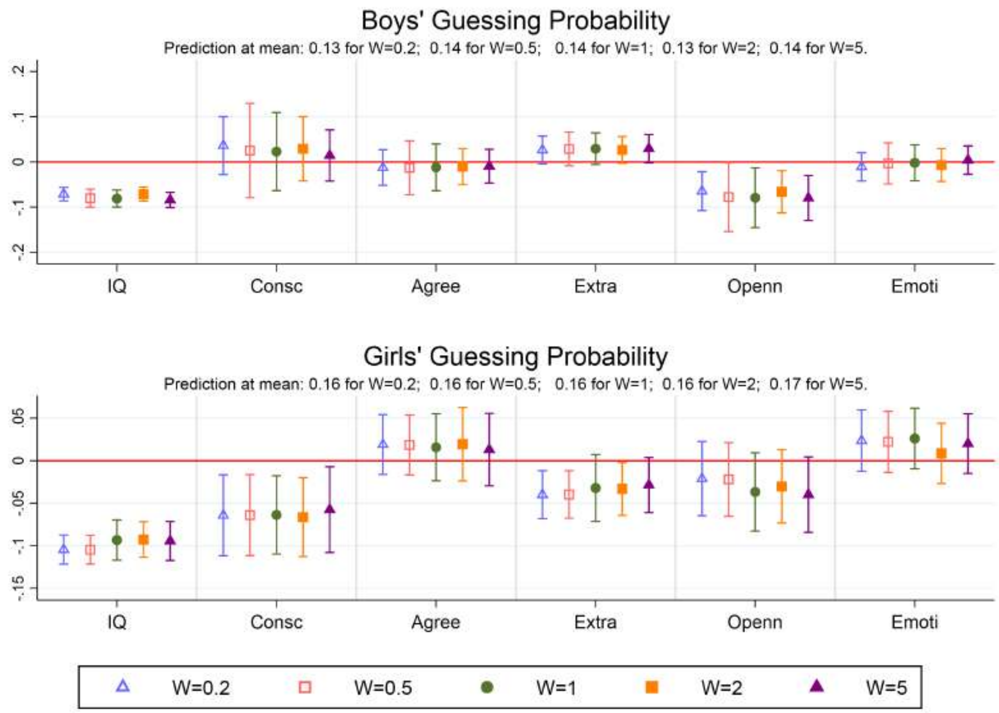


Figure M11: Guessing Probability Across Fallback Consumption Levels in CRRA

Notes: Effects of IQ or personality traits on guessing probability for boys (top) and girls (bottom) under the CRRA specification. Blue triangles represent $w_0 = 0.2$, red squares $w_0 = 0.5$, green circles $w_0 = 1$, orange squares $w_0 = 2$, and purple triangles $w_0 = 5$. Predictions at mean: boys = 0.13, 0.14, 0.14, 0.13, 0.14, girls = 0.16, 0.16, 0.16, 0.16, 0.17.

N Appendix N: Reference Wealth and Socioeconomic Background

This appendix examines whether socioeconomic status affects the reference point or fallback consumption 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, both parents, or voluntary migration).

For each utility specification, we present coefficient plots showing how cognitive skills (IQ) and non-cognitive skills (Big Five personality traits: 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 N1–N7 present results for the Expo-Power specification. Figures N8–N13 present corresponding results for the CRRA specification.

N.1 Expo-Power Specification

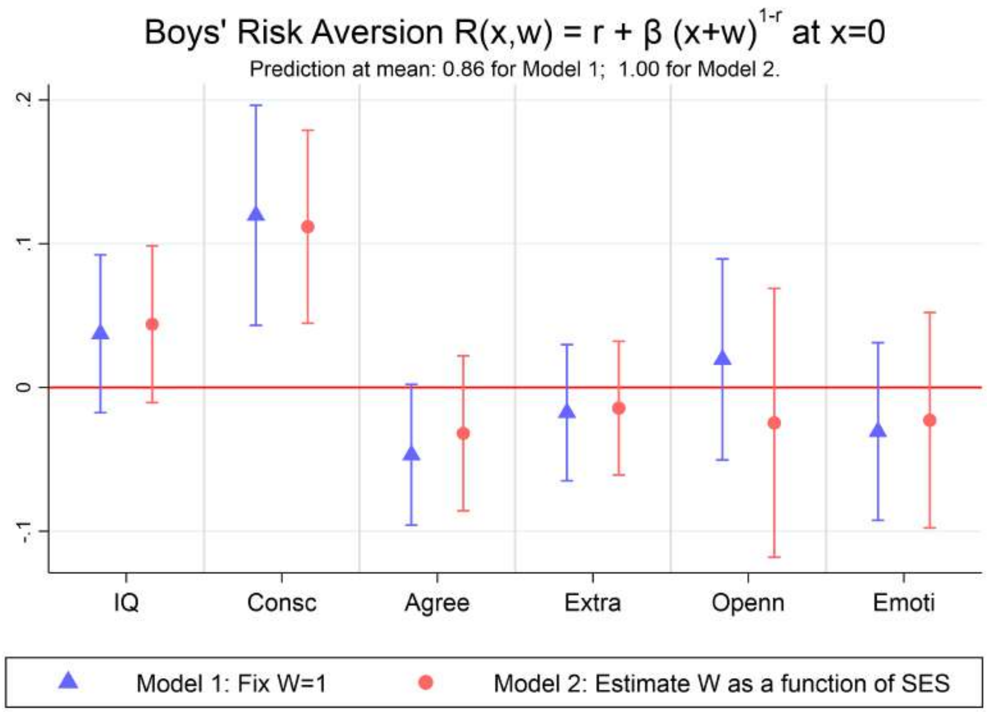


Figure N1: Risk Aversion: Boys

Notes: Effects of IQ and Big Five personality traits (Conscientiousness, Agreeableness, Extraversion, Openness, Emotional Stability) on risk aversion $R(x, w) = r + \beta(x + w)^{1-r}$ evaluated at $x = 0$ and $w = 1$ for boys. Blue squares represent Model 1 (fixed $W = 1$), red circles represent Model 2 (estimated W from SES). Predictions at mean: 0.86 for Model 1, 0.99 for Model 2.

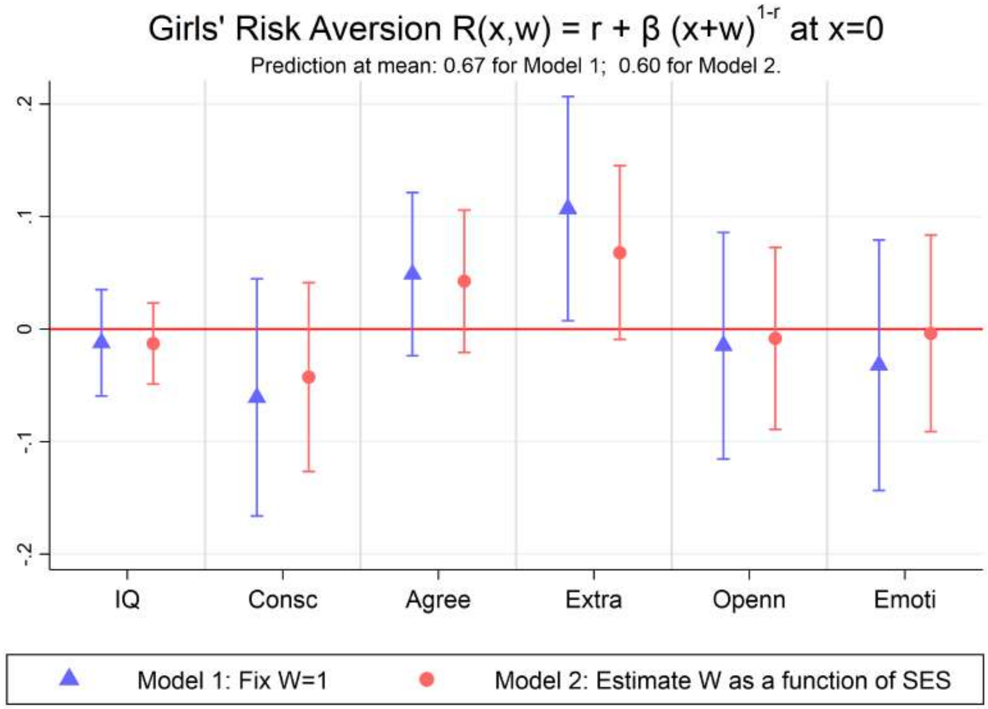


Figure N2: Risk Aversion: Girls

Notes: Effects of IQ and Big Five personality traits on risk aversion $R(x, w) = r + \beta(x + w)^{1-r}$ evaluated at $x = 0$ and $w = 1$ for girls. Blue squares represent Model 1 (fixed $W = 1$), red circles represent Model 2 (estimated W from SES). Predictions at mean: 0.67 for Model 1, 0.60 for Model 2.

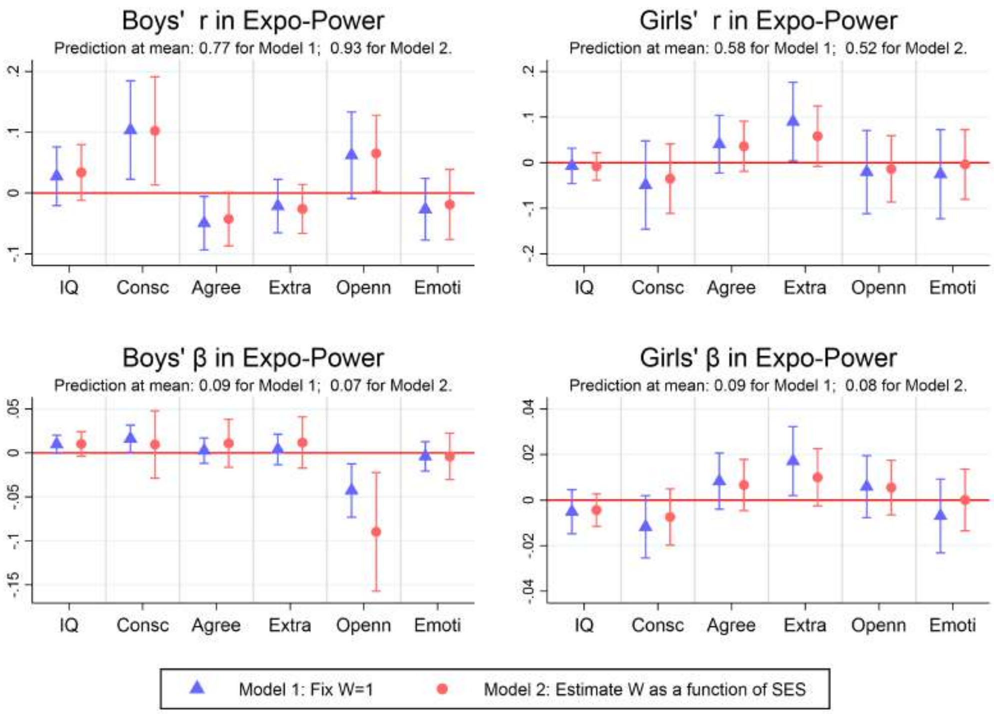


Figure N3: Expo-Power Parameters r and β

Notes: Effects of IQ and personality traits on Expo-Power parameters. Top panels show baseline curvature parameter r for boys (left) and girls (right). Bottom panels show wealth-sensitivity parameter β for boys (left) and girls (right). Blue squares represent Model 1 (fixed $W = 1$), red circles represent Model 2 (estimated W from SES).

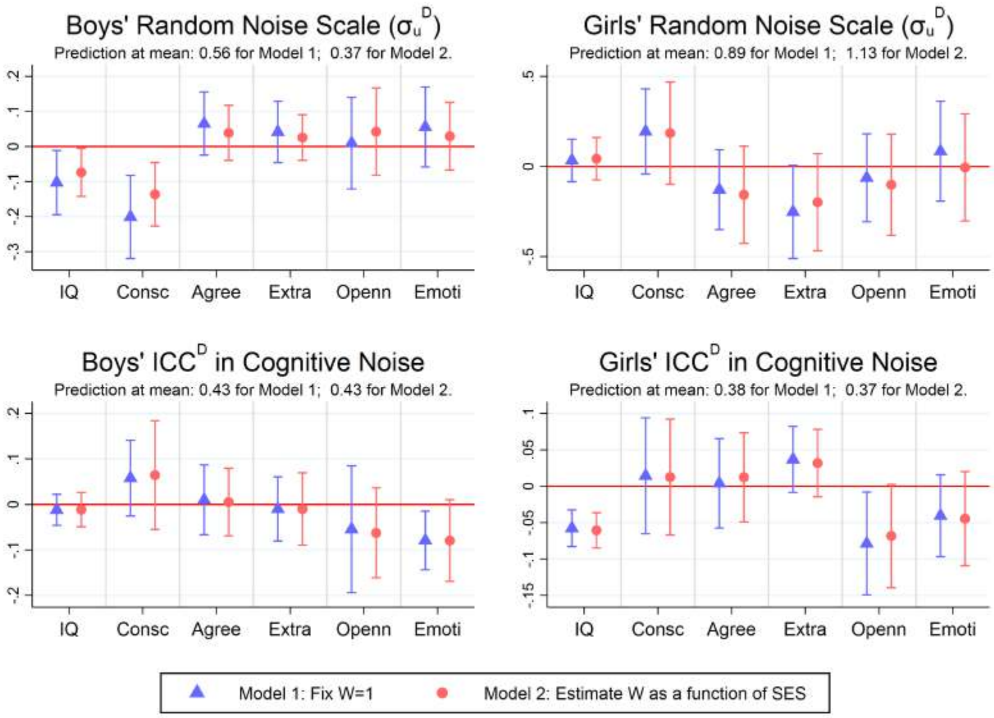


Figure N4: Random Noise Scale and ICC^D

Notes: Effects of IQ and personality traits on deliberation noise parameters. Top panels show random noise scale σ_u^D for boys (left) and girls (right). Bottom panels show ICC^D for boys (left) and girls (right). Blue squares represent Model 1 (fixed $W = 1$), red circles represent Model 2 (estimated W from SES).

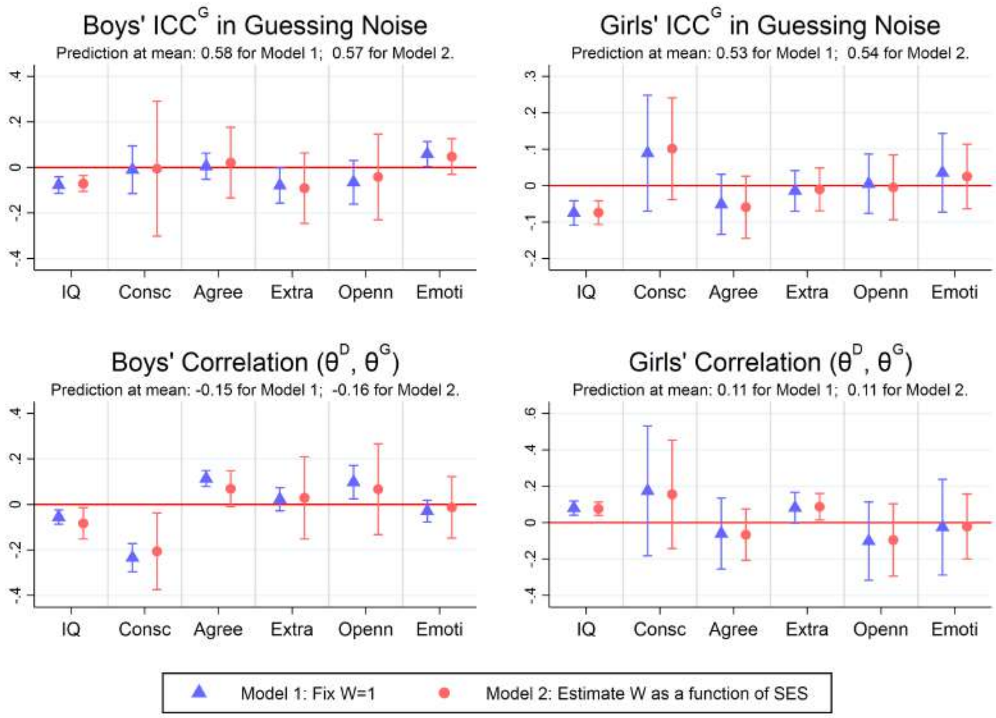


Figure N5: ICC^G and Correlation Between Random Effects

Notes: Effects of IQ and personality traits on guessing structure. Top panels show ICC^G for boys (left) and girls (right). Bottom panels show correlation between deliberation and guessing random effects $\rho(\theta^D, \theta^G)$ for boys (left) and girls (right). Blue squares represent Model 1 (fixed $W = 1$), red circles represent Model 2 (estimated W from SES).

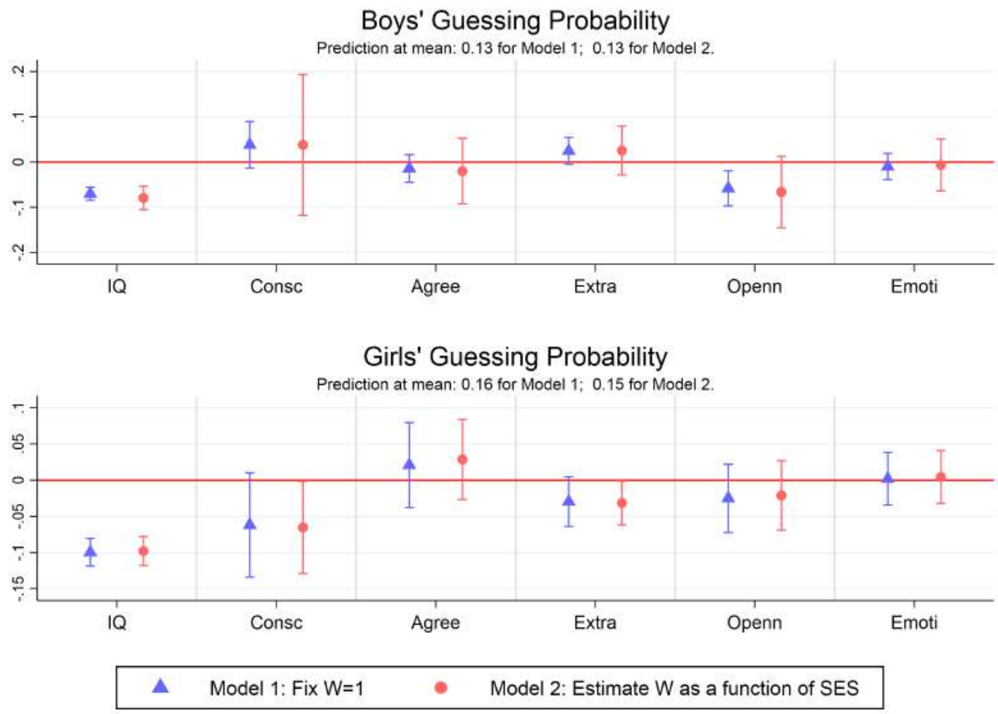


Figure N6: Guessing Probability

Notes: Effects of IQ and personality traits on guessing probability. Top panel shows boys, bottom panel shows girls. Blue squares represent Model 1 (fixed $W = 1$), red circles represent Model 2 (estimated W from SES). Predictions at mean: 0.13 for boys (both models), 0.16 for girls (both models).

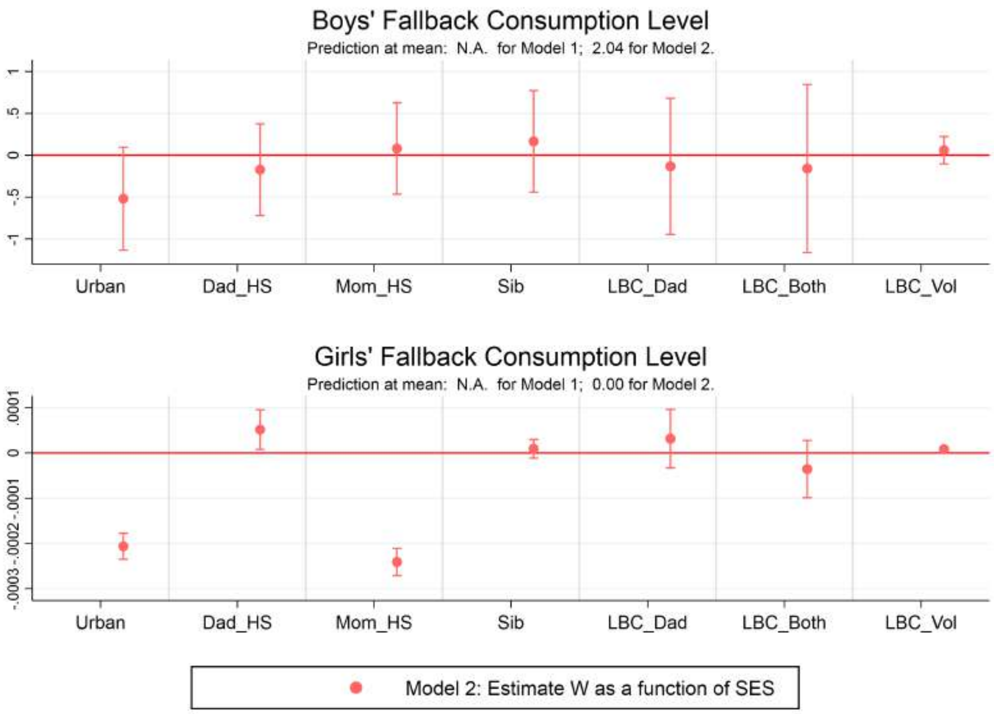


Figure N7: Fallback Consumption Across SES Variables

Notes: Effects of SES variables (urban residence, parental education, siblings, left-behind status) on estimated reference/fallback consumption W in Model 2. Top panel shows boys, bottom panel shows girls. Predicted fallback consumption at mean SES: 2.11 for boys, 0.00 for girls.

N.2 CRRA Specification

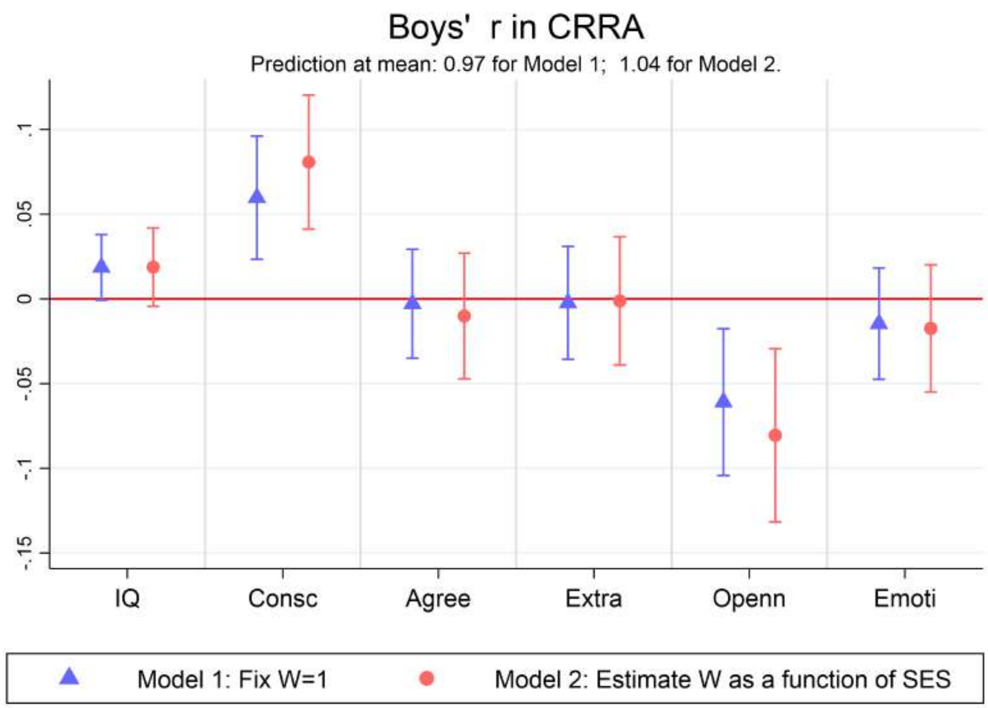


Figure N8: Risk Aversion Parameter in CRRA: Boys

Notes: Effects of IQ and Big Five personality traits (Conscientiousness, Agreeableness, Extraversion, Openness, Emotional Stability) on the risk aversion parameter r in the CRRA utility specification for boys. Blue squares represent Model 1 (fixed $W = 1$), red circles represent Model 2 (estimated W from SES). Predicted risk aversion at mean: 0.97 for Model 1, 1.04 for Model 2.

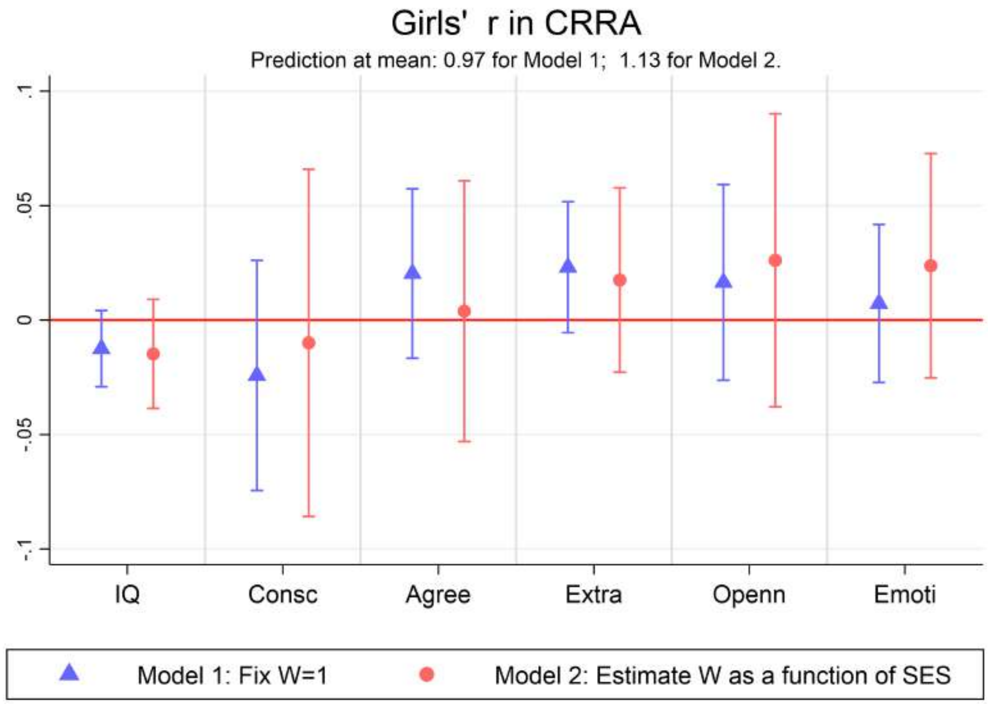


Figure N9: Risk Aversion Parameter in CRRA: Girls

Notes: Effects of IQ and Big Five personality traits (Conscientiousness, Agreeableness, Extraversion, Openness, Emotional Stability) on the risk aversion parameter r in the CRRA utility specification for girls. Blue squares represent Model 1 (fixed $W = 1$), red circles represent Model 2 (estimated W from SES). Predicted risk aversion at mean: 0.97 for Model 1, 1.13 for Model 2.

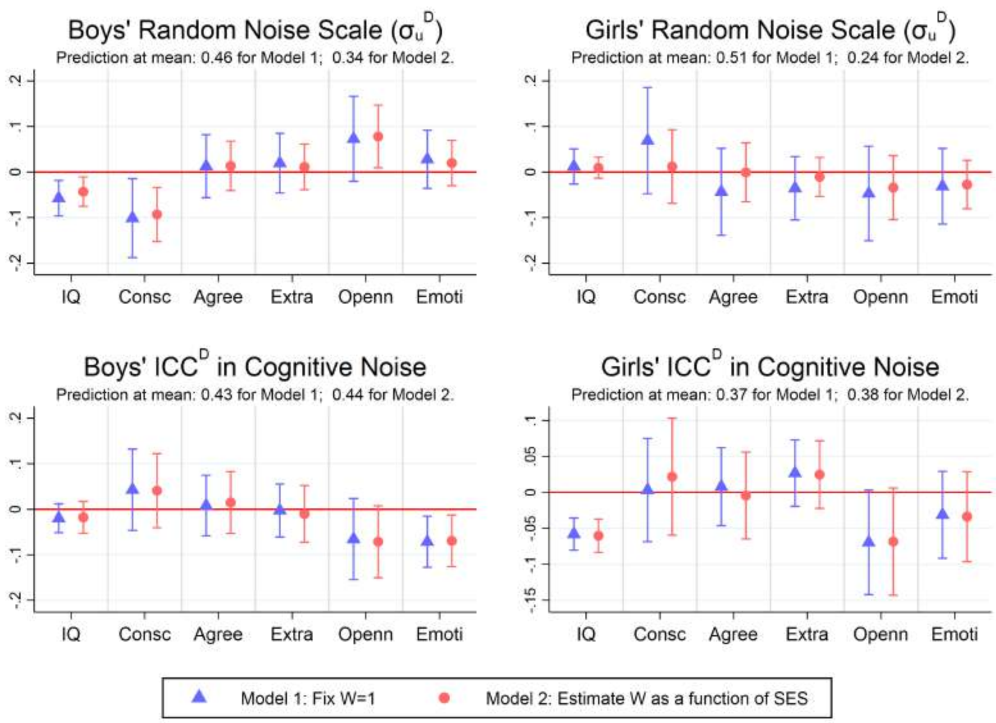


Figure N10: Error Structure Parameters in CRRA

Notes: Effects of IQ and Big Five personality traits (Conscientiousness, Agreeableness, Extraversion, Openness, Emotional Stability) on error structure parameters in the CRRA utility specification. Top panels show random noise scale (σ_u^D), bottom panels show ICC^D. Left panels show boys, right panels show girls. Blue squares represent Model 1 (fixed $W = 1$), red circles represent Model 2 (estimated W from SES). Predicted random noise scale at mean: 0.46 (Model 1) and 0.34 (Model 2) for boys, 0.51 (Model 1) and 0.24 (Model 2) for girls. Predicted ICC^D at mean: 0.43 (Model 1) and 0.44 (Model 2) for boys, 0.37 (Model 1) and 0.38 (Model 2) for girls.

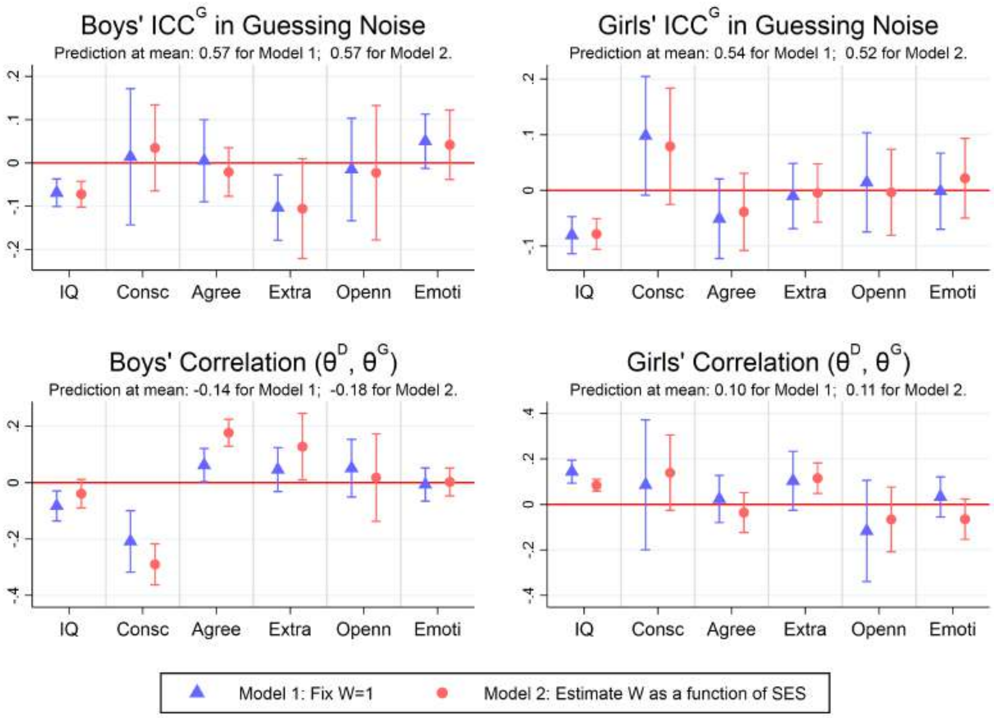


Figure N11: Additional Error Structure Parameters in CRRA

Notes: Effects of IQ and Big Five personality traits (Conscientiousness, Agreeableness, Extraversion, Openness, Emotional Stability) on additional error structure parameters in the CRRA utility specification. Top panels show ICC^G, bottom panels show correlation $\rho(\theta^D, \theta^G)$. Left panels show boys, right panels show girls. Blue squares represent Model 1 (fixed $W = 1$), red circles represent Model 2 (estimated W from SES). Predicted ICC^G at mean: 0.57 (Model 1) and 0.57 (Model 2) for boys, 0.54 (Model 1) and 0.53 (Model 2) for girls. Predicted correlation at mean: -0.14 (Model 1) and -0.15 (Model 2) for boys, 0.10 (Model 1) and 0.11 (Model 2) for girls.

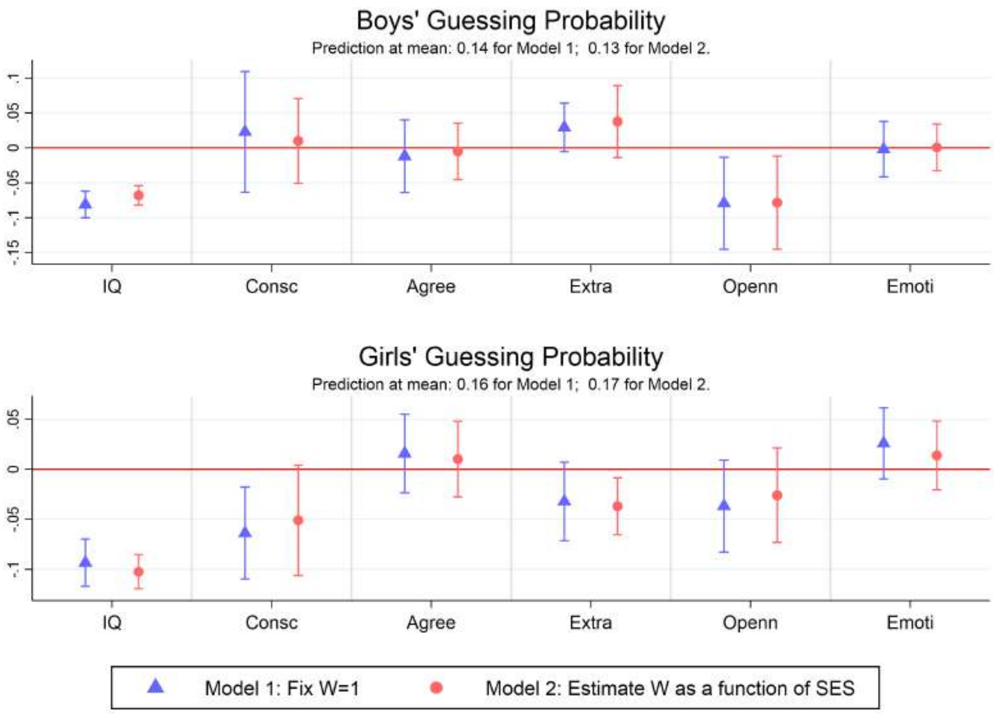


Figure N12: Guessing Probability in CRRA

Notes: Effects of IQ and Big Five personality traits (Conscientiousness, Agreeableness, Extraversion, Openness, Emotional Stability) on guessing probability in the CRRA utility specification. Top panel shows boys, bottom panel shows girls. Blue squares represent Model 1 (fixed $W = 1$), red circles represent Model 2 (estimated W from SES). Predicted guessing probability at mean: 0.14 (Model 1) and 0.13 (Model 2) for boys, 0.16 (Model 1) and 0.17 (Model 2) for girls.

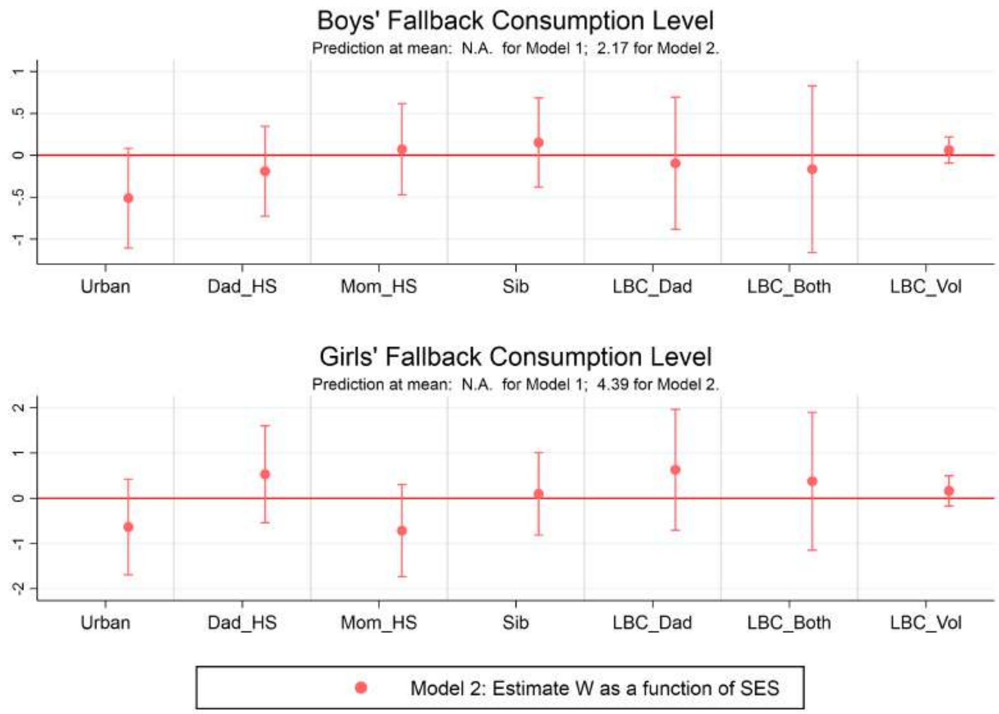


Figure N13: Fallback Consumption Across SES Variables in CRRA

Notes: Effects of SES variables (urban residence, parental education, siblings, left-behind status) on estimated reference/fallback consumption W in Model 2 for the CRRA utility specification. Top panel shows boys, bottom panel shows girls. Predicted fallback consumption at mean SES: 2.16 for boys, 4.38 for girls.

O Appendix O: Component Interactions and Superadditivity

This appendix documents the superadditive interactions between model components. Because superadditivity 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.

O.0.1 Component Interactions

Testing whether the various components' contributions are additive reveals substantial positive interaction. For boys' Expo-Power, components estimated separately yield 2,377.0 points (IQ + Personality: 383.7, ICC: 1,963.6, Guessing: 28.7, β parameter: 1.0), yet the full model improvement is 3,259.7 points. The 882.7-point difference represents 37% more explanatory power than simple additive contributions. For girls, superadditivity is 538.4 points (25% beyond the 2,164.4-point sum), as shown in Figure O1.

The β parameter contribution reveals striking 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 explains why functional form restrictions obscure girls' true preferences while barely affecting boys': CRRA's constant-curvature assumption binds girls' preferences tightly but leaves boys' essentially unconstrained. Combined with opposite error correlation patterns, this functional form sensitivity drives the gender gap reversal.

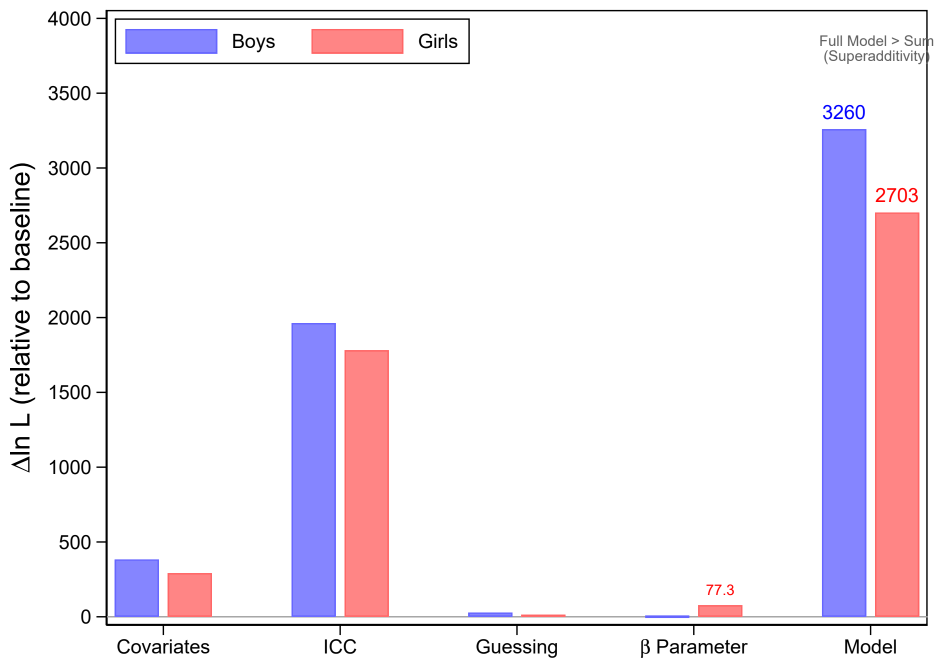


Figure O1: Component Contributions to Log-Likelihood Improvements

Notes: Log-likelihood improvements relative to baseline CRRA for boys (blue) and girls (red). First four bars show improvements when each component is added separately. Fifth bar shows actual improvement when all components are included jointly. Superadditivity—gap between sum of individual contributions and Full Model—demonstrates components interact positively: 883 points (37%) for boys, 538 points (25%) for girls. Boys: $N = 1{,}115$, Girls: $N = 1{,}089$.

The positive superadditivity indicates 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 decision noise.

The 77-fold difference in β parameter contribution between boys (1.0 points) and girls (77.3 points) noted in the main text reflects the greater importance of wealth-dependent risk aversion for girls. However, this difference in the β contribution does not directly explain the gender gap reversal, which operates primarily through the baseline curvature parameter r.

The superadditivity 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, decision precision, and strategic engagement—partial models that include only some components will systematically misestimate all parameters.