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

DISCUSSION PAPER SERIES

IZA DP No. 17168

How Cognitive Skills Affect Strategic Behavior: Cognitive Ability, Fluid Intelligence and Judgment

David Gill
Zachary Knepper
Victoria Prowse
Junya Zhou

JULY 2024

DISCUSSION PAPER SERIES

IZA DP No. 17168

How Cognitive Skills Affect Strategic Behavior: Cognitive Ability, Fluid Intelligence and Judgment

David Gill
Purdue University and IZA

Victoria Prowse
Purdue University and IZA

Zachary Knepper
Purdue University

Junya Zhou
UT-Dallas

JULY 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

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

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

www.iza.org

ABSTRACT

How Cognitive Skills Affect Strategic Behavior: Cognitive Ability, Fluid Intelligence and Judgment*

We explore the influence of cognitive ability and judgment on strategic behavior in the beauty contest game (where the Nash equilibrium action is zero). Using the level-k model of bounded rationality, cognitive ability and judgment both predict higher level strategic thinking. However, individuals with better judgment choose zero less frequently, and we uncover a novel dynamic mechanism that sheds light on this pattern. Taken together, our results indicate that fluid (i.e., analytical) intelligence is a primary driver of strategic level-k thinking, while facets of judgment that are distinct from fluid intelligence drive the lower inclination of high judgment individuals to choose zero.

JEL Classification: C92, C72, D91

Keywords: cognitive ability, judgment, fluid intelligence, matrix reasoning, beauty contest, strategic sophistication, level-k, experiment, game theory

Corresponding author:
David Gill
Department of Economics
Purdue University
Rawls 4023
Krannert School of Management
403 W. State Street
West Lafayette, IN 47907
USA

E-mail: dgill.econ@gmail.com

* We are grateful for helpful comments from seminar participants and in private conversations. We thank the Vernon Smith Experimental Economics Laboratory at Purdue University for hosting our experiment. We also thank the Krannert School of Management at Purdue University for funding our experiment. Finally, we thank Qilu Pan and Zhongheng Qiao for excellent research assistance.

1 Introduction

Cognitive skills matter for economic behavior and outcomes. For instance, Fe et al. (2022) show that cognitive ability measured in childhood predicts life outcomes, including educational attainment, fertility and success in the labor market. With increasing automation, higher-order cognitive skills are growing in importance (Deming, 2022). Alongside cognitive ability, these critical higher-order cognitive skills include: judgment (Chan and Schmitt, 2002), decision-making ability (Deming, 2021), problem solving (Autor, 2015), theory of mind (Fe et al., 2022) and creativity (Gill and Prowse, forthcoming).

In this paper we focus on the role that cognitive skills play in explaining strategic behavior. Behavior in strategic settings matters because people constantly engage in strategic interactions with others (Fe et al., 2022). For example, strategic sophistication helps people to cooperate successfully when building longterm relationships or when working together in teams, and strategic sophistication also helps people when they negotiate with others or compete with others in the labor market.

We focus on two cognitive skills: cognitive ability and judgment. To clarify our terminology, we include cognitive ability and judgment within the broader class of cognitive skills. As described by Fe et al. (2022), tests of cognitive ability "measure fluid intelligence (logical reasoning and problem-solving ability) or crystallized intelligence (acquired knowledge and verbal skills)." Like us, Fe et al. (2022) use a test of matrix reasoning to measure cognitive ability, and they "highlight the importance of theory of mind as a cognitive skill that is distinct from cognitive ability." Similarly, Gill and Prowse (forthcoming) describe a set of key "cognitive skills" that includes cognitive ability and creativity. We use a test of situational judgment that measures judgment in work-related situations. Our terminology follows the literature on situational judgment that views cognitive ability and judgment as related but distinct concepts (e.g., Weekley and Jones, 1997, 1999, Clevenger et al., 2001, and Weekley and Ployhart, 2005), with practical intelligence and interpersonal skills helping to underpin situational judgment (Stemler and Sternberg, 2006). Sections 2.4 and 2.5 provide a detailed discussion of what our tests of cognitive ability and judgment measure, and Section 3.2 reports a modest correlation of 0.34 between the cognitive ability and judgment test scores, which provides evidence that the tests capture different aspects of cognitive skills.

To study strategic behavior we use the p-beauty contest game (Nagel, 1995; Nagel et al., 2017). In our experiment, subjects repeatedly played the beauty contest in fixed groups of three. In each round, the three subjects in a group each chose a number from $\{0, ..., 100\}$, and the subject whose choice was closest to 70% of the mean of the three choices won a monetary prize. The beauty contest is well suited to study strategic behavior: although the incentive to undercut the average choice gives a unique Nash equilibrium at zero, behavior tends to the equilibrium slowly, and on the learning path choosing zero is often unprofitable; therefore, strategic skill in the beauty contest requires subjects to form accurate expectations about how others will choose and to best respond to those expectations. The beauty contest captures well the essence of many real-world strategic interactions, including those that exhibit unraveling such as the timing of asset sales during financial bubbles (Ho et al., 1998) or the timing of offers in labor market hiring (Gill and Prowse, 2023).

An influential stream of research finds that cognitive ability predicts strategic behavior in the p-beauty contest game: Burnham et al. (2009), Carpenter et al. (2013) and Fehr and Huck (2016) find that higher cognitive ability predicts lower choices, while Brañas-Garza et al. (2012), Gill and Prowse (2016) and Alós-Ferrer and Buckenmaier (2021) further use the level-k model of boundedly rational thinking to show that cognitive ability predicts strategic sophistication in the beauty contest (see also: Schnusenberg and Gallo, 2011; Capra, 2019; Taylor, 2020; and Hermes and Schunk, 2022).[1]

Cognitive ability is underpinned by fluid intelligence (also called analytical intelligence), which is "the ability to reason and solve problems involving new information, without relying extensively on an explicit base of declarative knowledge" (Carpenter et al., 1990). Fluid/analytical intelligence helps people to reason logically and to systemize new information. In strategic interactions such as the beauty contest, fluid/analytical intelligence bolsters strategic skill by helping people to understand the strategic context in which they and others are choosing, such as the rules of the strategic game and the payoff consequences of actions (Fe et al., 2022).

The robust finding in the literature that cognitive ability predicts strategic behavior in the p-beauty contest game leaves open two important related questions:

[1] Building on Ballester et al. (2021), Ballester et al. (2023) extend this literature by allowing beliefs about how level-0 subjects behave to vary across subjects: Ballester et al. (2023) find no evidence that cognitive ability predicts these beliefs.

- First, do cognitive skills other than cognitive ability also predict strategic behavior in the beauty contest?
- Second, to what extent is the predictive power of cognitive ability measuring the effect of fluid intelligence on strategic behavior versus the effect of other cognitive skills that are correlated with fluid intelligence?

These two questions are particularly pertinent because measures of fluid intelligence like matrix reasoning tend to correlate strongly with general intelligence g (see Gray and Thompson, 2004, Box 1).

To help answer these two questions, we study the extent to which cognitive ability and judgment predict strategic behavior in the p-beauty contest game. To measure cognitive ability, subjects completed a test of matrix reasoning that measures fluid intelligence and is similar to Raven's Progressive Matrices (Raven et al., 2000). To measure judgment, subjects completed a test of situational judgment that measures judgment in work-related situations. Practical intelligence and interpersonal skills help to underpin situational judgment ability (Stemler and Sternberg, 2006), and situational judgment scores predict job performance, even after controlling for cognitive ability, personality and experience (e.g., Clevenger et al., 2001). Sections 2.4 and 2.5 provide a detailed discussion of the tests.

We focus on judgment for three reasons. First, as we note above, judgment is one of a key set of higher-order cognitive skills that are increasingly important for success in life. Second, we are not aware of any existing work that studies whether judgment predicts behavior in the p-beauty contest game or other related strategic games. Third, the literature has conjectured that success in the beauty contest requires judgment: "The beauty-contest game exhibits characteristics of both an intellective task and a judgmental task. The judgmental aspect arises from the interactive structure of the beauty-contest game. Correct expectations on other decision makers' guesses are crucial for one's own guess. Obviously, the intellective task consists in the iterated elimination of dominated strategies and the correct adaptation of one's own guess to one's expectations of guesses submitted by other participants" (Kocher and Sutter, 2005).

As described above, our first question is whether cognitive skills other than cognitive ability also predict strategic behavior in the p-beauty contest game. We find evidence that both cognitive ability and judgment predict strategic behavior. Following, e.g., Nagel (1995), Duffy and Nagel (1997), Ho et al. (1998), Gill and Prowse (2016) and Alós-Ferrer and Buckenmaier (2021), we use the lens of the level-k model of boundedly rational thinking to study strategic sophistication in the beauty contest. In the level-k model, level-0 types are strategically unsophisticated, level-1 types best respond to all others being level-0, level-2 types best respond to all others being level-1, and so on. Based on estimates from a structural level-k model that allows learning across rounds, we find that judgment and cognitive ability both predict level-k thinking. Specifically, better judgment and higher cognitive ability predict higher level-k on average, with the effect of judgment on level-k thinking about half the size of that of cognitive ability.

Even though cognitive ability and judgment predict level-k thinking in the same direction, we also find interesting differences in the pattern of behavior. In particular, subjects who are higher in cognitive ability are more likely to choose zero, while subjects who are higher in judgment are *less likely* to choose zero.[2] We uncover a novel dynamic mechanism that sheds light on this pattern. Specifically, we study how the probability that a subject chooses zero in a round increases as their group's average choice in the previous round falls toward zero. We find that for subjects who are high in judgment, this probability of choosing zero responds less strongly to the group's average choice in the previous round, compared to subjects who are low in judgment.

As described above, our second question is the extent to which the predictive power of cognitive ability measures the effect of fluid intelligence on strategic behavior versus the effect of other cognitive skills that are correlated with fluid intelligence. When we simultaneously include judgment and cognitive ability in our level-k model, we find that the coefficient on cognitive ability is stable and remains statistically significant, while the coefficient on judgment falls and becomes statistically insignificant. The stability of the coefficient on cognitive ability supports the hypothesis that the power of cognitive ability to predict strategic behavior is driven by the fluid/analytical intelligence that underpins cognitive ability. At the same time, the instability of the coefficient on judgment provides evidence that the power of judgment to predict level-k thinking is driven mostly by fluid intelligence that helps judgment, with other important facets of judgment (including

[2]Although zero is the Nash equilibrium action, subjects might choose zero for a variety of reasons. Therefore, at the suggestion of a helpful referee, when describing our results we do not mention the word "equilibrium." To delve further into this point, future work could replicate our design using an interior Nash equilibrium.

practical intelligence and interpersonal skills) playing a lesser role. Taken together, these results indicate that fluid intelligence is a primary driver of strategic level-k thinking.

Interestingly, the result described above that subjects who have better judgment are less likely to choose zero holds whether or not we include cognitive ability in our reduced-form regressions. This provides evidence that the lower inclination of high judgment subjects to choose zero is driven by facets of judgment that are distinct from fluid intelligence.

The paper proceeds as follows: Section 2 describes the experimental design, Section 3 provides the main results, Section 4 presents further analyses, and Section 5 concludes. The Web Appendix provides further details.

2 Experimental design

2.1 Subject pool and procedures

We collected experimental data from 141 student subjects at the Vernon Smith Experimental Economics Laboratory (VSEEL) at Purdue University during November 2021 (AEARCTR-0008497; Purdue IRB-2021-1558). We drew subjects from the VSEEL subject pool (administered using ORSEE; Greiner, 2015), and the experiment was programmed in oTree (Chen et al., 2016). Subjects received a $5 show-up fee. Web Appendix I.1 provides the experimental instructions.

2.2 Beauty contest game

Subjects repeatedly played the p-beauty contest game, with $p = 0.7$, in fixed groups of three (i.e., with no rematching). In each round, each of the three subjects in a group simultaneously chose an integer $x_i \in \{0, ..., 100\}$, and the subject whose choice was closest to 70% of the mean of the three choices won a prize of $6 (in the case of ties, the prize was split equally among the tied subjects). At the end of each round, subjects received feedback about: the group's choices in that round; 70% of the mean of the choices; which group member (or members) was closest to 70% of the mean; and the earnings of each group member in that round.[3] In our parameterization, the unique Nash equilibrium is for all three group members to choose zero. Each group of three subjects played the game between five and ten times (with random termination starting from the end of round 5). In total, we collected 999 subject-round observations. Web Appendix I.2 provides further details.

As we explain in Web Appendix I.2, Online Appendix II.6 of Gill and Prowse (2023) also uses the same dataset to show that subjects are not forward-looking in the p-beauty contest game; however, the cognitive ability and judgment measures that are central to this paper are not analyzed there.

Finally, we note that the three-person beauty contest game is more complex than the large-n beauty contest in the sense that a subject's own choice has a non-negligible impact on the mean, which makes calculating best responses to beliefs more difficult.

[3]Each group member was identified using a label (X, Y, or Z) that was constant across rounds. Feedback in our experiment is similar to Nagel (1995) who revealed the whole distribution of choices at the end of each round.

Indeed, based on data from the three-person beauty contest game, Gill and Prowse (2016) find that their structural model fits better when the model assumes that individuals fail to take into account the effect of their own choice on the target. Relatedly, using the same dataset as Gill and Prowse (2016), Gill and Prowse (2023) find that the most complex situation (measured using response times) is the one in which a subject won in the previous round with a choice between that of her two opponents: one interpretation is that this perhaps unexpected outcome forces the subject to pay more attention to the complexity of three-person beauty contest best-response calculations. As described in detail by Fe et al. (2022, Section IV.A) in the context of the level-k model, calculating best responses in undercutting games mainly uses cognitive ability, while forming beliefs about opponents' behavior mainly uses first and higher-order theory-of-mind ability. Therefore, the complexity of the three-person beauty contest might influence the importance of cognitive skills.

2.3 Tests of cognitive skills and demographic questionnaire

After the p-beauty contest, subjects completed two tests of cognitive skill from the International Cognitive Ability Resource (ICAR), a joint venture between researchers at Northwestern and the University of Cambridge, among others, that provides a suite of public-domain tests (Dworak et al., 2021; `https://icar-project.com`). To measure judgment, subjects first completed an eleven-item test of situational judgment. This test measures judgment in work-related situations, and Section 2.5 discusses the test. To measure cognitive ability, subjects then completed an eleven-item test of matrix reasoning similar to Raven's Progressive Matrices (Raven et al., 2000). This test measures fluid/analytical intelligence, and Section 2.4 discusses the test. Finally, subjects completed a short demographic questionnaire. Web Appendix I.3 provides further details.

2.4 Discussion of the cognitive ability test

To measure cognitive ability, subjects completed a test of matrix reasoning that measures fluid/analytical intelligence. Web Appendix I.3 provides details of the test and Web Appendix I.6 provides one of the items as a sample. This test of matrix reasoning requires subjects to identify missing elements that complete visual patterns presented using matrices, and is similar to Raven's Progressive Matrices (Raven et al., 2000).

As described by Fe et al. (2022), tests of cognitive ability "measure fluid intelligence (logical reasoning and problem-solving ability) or crystallized intelligence (acquired knowledge and verbal skills)." Matrix reasoning is recognized as a leading measure of fluid intelligence (Carpenter et al., 1990, Gray and Thompson, 2004). For example, Gray and Thompson (2004, Box 1) illustrates many different tests of fluid and crystallized intelligence, describes how "scores on various tests can be factor-analysed to give g, a single summary measure of cognitive ability," and depicts how matrix reasoning has the strongest correlation with g. Similarly, according to Carpenter et al. (1990), matrix reasoning captures abstract reasoning and correlates strongly with performance on other complex cognitive tasks.

In economics, scores on matrix reasoning tests have been found to predict the ability to form and update beliefs accurately (Burks et al., 2009, Charness et al., 2018), and matrix reasoning tests have been used in a variety of games to study the effect of cognitive ability on strategic behavior by, e.g., Gill and Prowse (2016), Proto et al. (2019, 2022), Fe et al. (2022), Hermes and Schunk (2022) and Gill and Rosokha (forthcoming).

2.5 Discussion of the judgment test

To measure judgment, subjects completed a test of situational judgment that measures judgment in work-related situations. Web Appendix I.3 provides details of the test, Web Appendix I.4 provides one of the items as a sample, and Web Appendix I.5 provides a brief description of the eleven situations used in our eleven-item test.

Situational judgment tests are "designed to measure judgment in work settings" (McDaniel and Nguyen, 2001), "have been used in personnel selection for about eighty years" (McDaniel et al., 2011), and consist of "challenging work-related situations and a list of plausible courses of action" (Oostrom et al., 2015). Specifically, situational judgment tests measure the ability to judge effective actions in different situations (Whetzel and McDaniel, 2009), which requires the ability to evaluate information, to relate effectively with others, and to adapt to changing circumstances (Clevenger et al., 2001). These different situations involve realistic practical problems that tend to be ill-defined, with incomplete information and multiple possible solutions of varying effectiveness, in contrast to academic problems that tend to be well-defined, with complete information and a single correct answer (Schmitt and Chan, 2006).

The power of situational judgment scores to predict job performance is partially explained by cognitive ability (e.g., Weekley and Jones, 1997). However, a substantial body of evidence finds that situational judgment scores predict job performance, even after controlling for cognitive ability, personality and experience (e.g., Weekley and Jones, 1997, 1999, Clevenger et al., 2001, and Weekley and Ployhart, 2005).[4]

Stemler and Sternberg (2006) emphasize that practical intelligence helps to underpin situational judgment.[5] As described by Sternberg et al. (1995), the hallmark of practical intelligence is the acquisition and use of "tacit" action-oriented knowledge, which takes the form of "knowing how" as opposed to "knowing that," and which is often difficult to articulate clearly. Sternberg (1999) notes that "practical intelligence as embodied in tacit knowledge increases with experience," and Sternberg et al. (1995) further note that "the ability to solve problems of a practical nature is maintained or even increased through late adulthood," while "the ability to solve strictly academic problems declines from early to late adulthood."[6]

Stemler and Sternberg (2006) further explain that the practical intelligence that underpins situational judgment shares some overlap with "social intelligence," while Lievens et al. (2008)'s review notes that, in part, situational judgment tests measure interpersonal skills. Developing this point, Chan and Schmitt (2002) find that situational judgment scores predict not only overall job performance, but also interpersonal performance (to measure interpersonal performance, supervisors assessed performance in "interpersonal conflict resolution, negotiation, and teamwork and cooperation").

[4]These four papers find mean correlations between situational judgment and cognitive ability of 0.28, 0.45, 0.27, and 0.36, respectively, while McDaniel et al. (2007)'s meta-study ($N = 30,859$) finds a correlation of 0.32. These correlations from the literature are similar to the correlation of 0.34 in our data (see Section 3.2). Furthermore, these papers and McDaniel et al. (2007)'s meta-study find correlations between situational judgment and personality or experience that are lower than those for cognitive ability.

[5]Stemler and Sternberg (2006) split intelligence into three aspects: creative, analytical, and practical.

[6]There is also a potential link between practical intelligence and judgment in the domain of creativity: Gill and Prowse (forthcoming) describe how creativity requires judgment to evaluate new ideas, and find that "creative individuals tend to have a package of practical skills that allows them to thrive in work environments in which learning from experience is important."

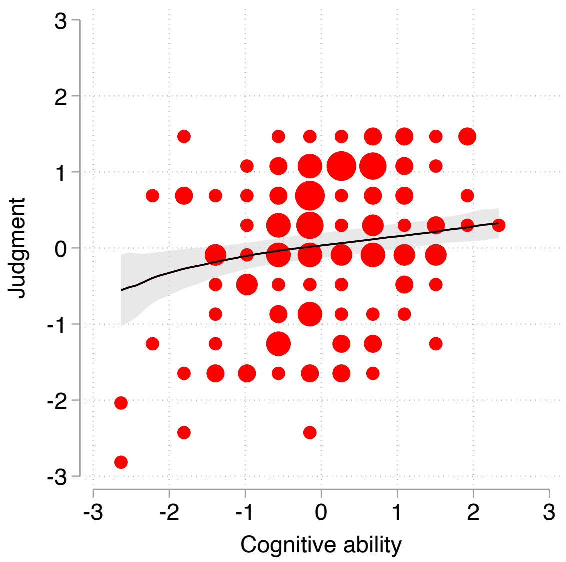
3 Main results

3.1 Time trend of choices

We confirm the finding from previous p-beauty contest game experiments that choices fall slowly over rounds toward zero. In our data, the mean choice across subjects falls from 43.9 in round 1 to 5.5 in round 10, with a linear round trend of -4.2 ($p < 0.001$, from an ordinary least squares (OLS) regression).

3.2 Test scores: standardization and correlation

Scores on the judgment test have a mean of 7.7 (out of 11) and a standard deviation of 1.2. Scores on the cognitive ability test have a mean of 7.2 (out of 11) and a standard deviation of 2.6. We standardize test scores to obtain judgment and cognitive ability scores that each have a mean of zero and standard deviation of one. The Pearson correlation between the judgment and cognitive ability scores is 0.34 ($p < 0.001$), which provides evidence that our two tests capture different aspects of cognitive skills.[7] Figure 1 provides a scatter plot.



Notes: Test scores are standardized. Red dots represent the data, with their area proportional to the number of subjects. The black line is the smoothed local polynomial fit of judgment on cognitive ability, and the grey area is the 95% confidence interval.

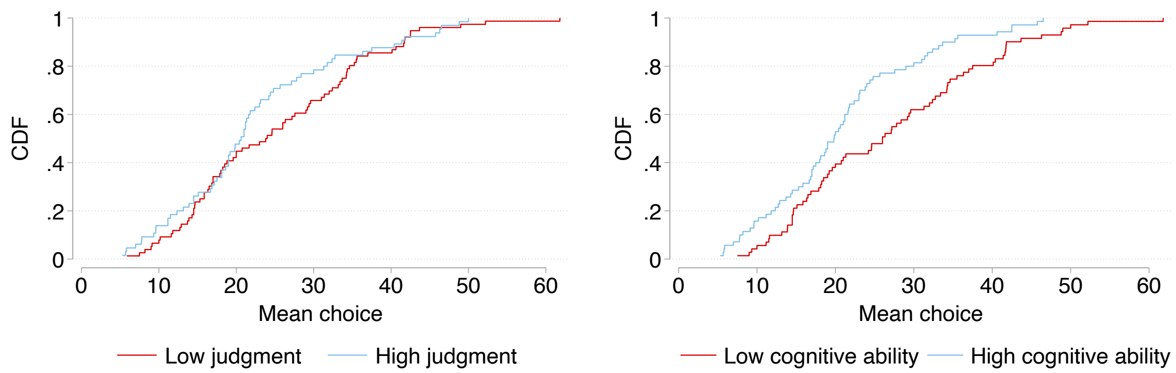
Figure 1: Skills scatter plot

[7]For comparison, Fe et al. (2022) find a correlation of 0.28 between cognitive ability (also measured using matrix reasoning) and theory of mind, while Gill and Prowse (forthcoming) find a correlation of 0.54 between cognitive ability and creativity.

3.3 Reduced-form results

As described in the introduction, our first question is whether cognitive skills other than cognitive ability also predict strategic behavior in the p-beauty contest game. In this section, we provide reduced-form evidence that both cognitive ability and judgment predict strategic behavior in the beauty contest.

We start by looking at mean choices across rounds at the subject level. Figure 2 uses empirical cumulative distribution functions (CDFs) to visualize the effects of cognitive skills on mean choices. The left panel shows the empirical CDFs split by high and low judgment subjects (where we split subjects into high/low categories using the median test score), while the right panel shows the empirical CDFs split by high and low cognitive ability subjects. We can see that both high judgment subjects and high cognitive ability subjects tend to choose lower numbers (that is, numbers closer to zero). The effects are strongest around the 75th percentile: at that percentile of mean choices, high judgment subjects' choices are around 6.6 lower than those of low judgment subjects ($p = 0.029$; two-sided test from an unconditional quantile regression with clustering at the group level), while high cognitive ability subjects' choices are around 11.3 lower than those of low cognitive ability subjects ($p = 0.004$). These effects of cognitive skills on mean choices translate into differences in earnings in the p-beauty contest game: Figure A.1 in Web Appendix II shows the empirical CDFs for earnings.



Notes: In the left panel, we classify each subject as 'high judgment' if their score on the judgment test was strictly above the sample median or 'low judgment' if their score was at or below the median. In the right panel, we classify in the same way, but using scores on the cognitive ability test.

Figure 2: CDFs of subject-level mean choice across rounds

From Figure 2, better judgment and higher cognitive ability both predict lower average choices. However, we also find interesting differences in the pattern of behavior by type of cognitive skill. Specifically, Table 1 provides evidence that subjects who are higher in cognitive ability are more likely to choose zero, while subjects who are higher in judgment are *less likely* to choose zero. Since choices fall over rounds toward zero (Section 3.1), the effects are bigger and statistically significant in the second half: when we include both cognitive skills in the same regression (specification (3) in Panel III), a one-standard-deviation increase in judgment predicts a five-percentage-point decrease in the proportion of choices of zero (two-sided $p = 0.009$), while a one-standard-deviation increase in cognitive ability predicts a five-percentage-point increase in the proportion of choices of zero (two-sided $p = 0.059$). These effect sizes are large when compared to the average proportion of choices of zero in the second half of 0.09. Table A.1 in Web Appendix II shows that we do not observe similar effects of cognitive skills on close-to-zero choices.

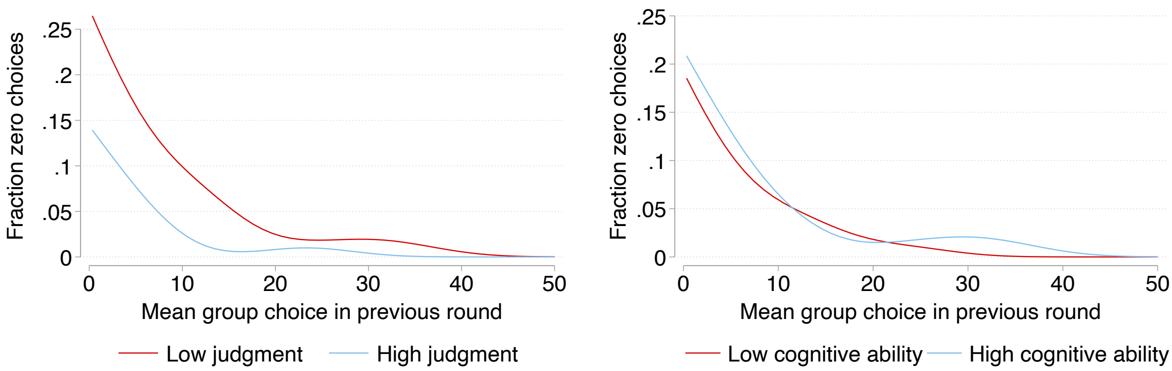
The result in Table 1 that subjects who are higher in judgment are less likely to choose zero holds whether or not we include cognitive ability in our regressions (in fact, the absolute magnitude of the effect is slightly larger when we include cognitive ability in specification (3)). Since our matrix reasoning test of cognitive ability measures fluid/analytical intelligence, this provides evidence that the lower inclination of high judgment subjects to choose zero is driven by facets of judgment that are distinct from fluid intelligence.

	Chose zero in round		
	(1)	(2)	(3)
Panel I: All rounds			
Effect of judgment (1 SD increase)	-0.015*		-0.024**
	(0.008)		(0.011)
Effect of cognitive ability (1 SD increase)		0.019*	0.026*
		(0.011)	(0.014)
Mean of dependent variable	0.05	0.05	0.05
Subject-round observations	999	999	999
Panel II: First half			
Effect of judgment (1 SD increase)	-0.005		-0.008
	(0.007)		(0.009)
Effect of cognitive ability (1 SD increase)		0.007	0.009
		(0.007)	(0.009)
Mean of dependent variable	0.02	0.02	0.02
Subject-round observations	564	564	564
Panel III: Second half			
Effect of judgment (1 SD increase)	-0.031**		-0.045***
	(0.014)		(0.017)
Effect of cognitive ability (1 SD increase)		0.038	0.051*
		(0.024)	(0.027)
Mean of dependent variable	0.09	0.09	0.09
Subject-round observations	435	435	435

Notes: Each specification reports estimates from an OLS regression that includes round indicators and controls for the demographics that we collected (see Web Appendix I.3). To give a more equal split of subject-round observations across halves, we define the first half as rounds 1-4 (recall random termination began from round 5). Since the test scores are standardized, each coefficient gives the effect of a one-standard-deviation (1 SD) increase. Heteroskedasticity-robust standard errors, clustered at the group level, are shown in parentheses. ***, **, and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table 1: Effects of cognitive skills on proportion of choices of zero

We uncover a novel dynamic mechanism that helps to shed light on why higher judgment subjects are less likely to choose zero. In particular, Figure 3 illustrates how the probability that a subject chooses zero in a round increases as their group's average choice in the previous round falls toward zero. From the left panel, we see that high judgment subjects' probability of choosing zero responds less strongly to their group's average choice in the previous round, compared to low judgment subjects. By contrast, the right panel shows no systematic difference by cognitive ability category.[8]



Notes: The panels show the relationship between the proportion of choices of zero in round r and the mean choice in the subject's group of three in round $r-1$, estimated using local polynomial regressions with a Gaussian kernel and using a bandwidth of 5. The notes to Figure 2 describe how we classify subjects by cognitive skill level. We excluded observations where the mean group choice in the previous round was zero (i.e., all three subjects chose zero) or strictly greater than 50.

Figure 3: Effect of mean group choice in previous round on proportion of choices of zero

[8]In order to accommodate more general games with interior Nash equilibria, Figure A.3 in Web Appendix II re-interprets the relationship described in Figure 3 as the relationship between the probability of choosing the equilibrium action and the group's mean distance to equilibrium in the previous round. Furthermore, in Figure A.3 we add structure by estimating a probability model where the probability of choosing the equilibrium action is a logistic function of the group's mean distance to equilibrium in the previous round (see the notes to Figure A.3 for details).

3.4 Structural level-k results

Our reduced-form results above provide evidence that judgment and cognitive ability both predict strategic behavior in the beauty contest game. To understand more about how cognitive skills influence strategic sophistication, we use the lens of the level-k model of boundedly rational thinking. In the level-k typology, level-0 types choose according to some strategically unsophisticated rule, level-1 types best respond to all others being level-0, level-2 types best respond to all others being level-1, and so on (Nagel, 1995; see also Stahl and Wilson, 1994, for a closely related model). Alaoui and Penta (2016, 2022)'s model of endogenous depth of reasoning, in which the number of steps of reasoning is determined by a comparison of the cost and value of each additional step, microfounds subjects' choice of level, while Alós-Ferrer and Buckenmaier (2021) and Gill and Prowse (2023) find that higher level-k types think for longer in the beauty contest.

We estimate the structural level-k model from Gill and Prowse (2016). Equation (6) in Section V.C describes how an individual's probability of being each level-k type, $Pr(k|\mathbf{z}_i)$, depends on a vector $\mathbf{z}_i$ that captures personal traits of the individual (e.g., cognitive ability and personality in the setting of Gill and Prowse, 2016). Specifically, letting K be the set of level-k types,

$$\Pr(k|\mathbf{z}_i) = \frac{\exp(\boldsymbol{\lambda}_k \mathbf{z}_i)}{\sum_{j \in K} \exp(\boldsymbol{\lambda}_j \mathbf{z}_i)}. \tag{1}$$

In Section V.C of Gill and Prowse (2016), $K = \{0, 1, 2, 3, 4\}$: due to limitations imposed by the smaller number of observations here, we reduce the number of parameters by removing the level-4 type.[9]

Note that we never allocate a particular subject in our experiment to a particular level-k type. Instead, we use Gill and Prowse (2016)'s standard mixture-of-types model, where an individual's personal traits $\mathbf{z}_i$ imply a probability distribution over the level-k types, given by equation (1). We calculate the level-k type probabilities in Panel I of Table 2 by averaging the estimated type probabilities across subjects (see the notes to Table 2 for details).

[9] As in Gill and Prowse (2016): (i) $\mathbf{z}_i$ includes an element that equals unity for all subjects, and therefore the model includes a full set of type-specific intercepts; and (ii) we impose the identifying normalization $\boldsymbol{\lambda}_1 = \mathbf{0}$.

In Gill and Prowse (2016)'s model, a level-k type follows the noisy level-k choice rule in every round $r \geq 2$ (round 1 choices seed the model). The model includes learning in order to capture the tendency of choices to move toward zero across rounds. The level-0 type learns in a strategically unsophisticated manner by "following the crowd" and copying their group's average choice from the previous round, while agents with level-$k > 0$ understand how lower-level agents learn across rounds.[10]

Specifically, Gill and Prowse (2016) assume that choices are independent draws from discretized and truncated t-distributions. Letting $f(\cdot|\mu, \sigma, \nu)$ be the density of the t-distribution with mean μ, scale σ and degrees of freedom ν, the probability of choice x in round $r \geq 2$ when following the level-k choice rule, given the group-specific mean in the previous round $\overline{x}_{r-1}$, is

$$\Pr(x|k, \overline{x}_{r-1}) = (1 - \gamma(r)) \frac{f(x|\mu, \sigma, \nu)}{\sum_{j=0}^{99} f(j|\mu, \sigma, \nu)} \mathbf{1}_{x \in \{0,1,\ldots,99\}} + \gamma(r) \mathbf{1}_{x=100} \tag{2}$$

where

$$\mu = \left(\frac{7}{10}\right)^k \overline{x}_{r-1}, \tag{3}$$

rounded to the nearest integer, and $\gamma(r)$ is the the empirical frequency of choices of 100 in our sample in round r.[11]

According to equation (3), individuals who follow the level-k choice rule for $k > 0$ do not take into account the effect of their own choice on the target; instead they aim to hit the target that would result from all group members noiselessly following the level-$(k-1)$ choice rule. Gill and Prowse (2016) find that an alternative specification in which individuals do take into account the effect of their own choice on the target fits their data less well.

Table 2 presents estimates from this structural level-k model. Panel I reports level-k type probabilities that match the pattern found previously in the literature (with level-1 and level-2 being the most common types; Crawford et al., 2013). Panel II reports the marginal effects of cognitive skills on the average level-k. Specifications (1) and (2) show that both judgment and cognitive ability predict level-k thinking. A one-standard-

[10]This level-k model of learning builds on Nagel (1995). See also Stahl (1996), Duffy and Nagel (1997), and Ho et al. (1998).

[11]Equation (3) in Gill and Prowse (2016) describes how the scale parameter σ depends on μ and r. Due to limitations imposed by the smaller number of observations here, we reduce the number of parameters by setting $\boldsymbol{\beta} = 0$, and so σ does not depend on μ.

deviation increase in judgment increases the average level-k by around 0.15, which is a 9% increase relative to the average (specification (1); two-sided $p = 0.090$).[12] A one-standard-deviation increase in cognitive ability increases the average level-k by around 0.29 (specification (2); two-sided $p < 0.001$).[13] Thus, the effect of judgment on level-k thinking is about half the size of that of cognitive ability.

When we include both judgment and cognitive ability in the same model (specification (3) in Table 2), the coefficient on cognitive ability is stable and remains statistically significant, while the coefficient on judgment falls and becomes statistically insignificant. These results provide evidence relevant to our second question. As described in the introduction, this second question is the extent to which the power of cognitive ability to predict strategic behavior measures the effect of fluid/analytical intelligence versus the effect of other cognitive skills that are correlated with fluid intelligence. Recalling that our matrix reasoning test of cognitive ability measures fluid intelligence, the stability of the coefficient on cognitive ability supports the hypothesis that the power of cognitive ability to predict level-k thinking is driven by the fluid intelligence that underpins cognitive ability. At the same, the instability of the coefficient on judgment provides evidence that the power of judgment to predict level-k thinking (specification (1)) is driven mostly by fluid intelligence that helps judgment, with other important facets of judgment (including practical intelligence and interpersonal skills) playing a lesser role.

To briefly summarize, the level-k analysis provides evidence that: (i) cognitive ability and judgment both predict level-k thinking; (ii) the power of cognitive ability to predict level-k thinking is driven by the fluid intelligence that underpins cognitive ability; and (iii) the power of judgment to predict level-k thinking is driven mostly by the fluid intelligence that helps judgment, with other important facets of judgment playing a lesser role. Taken together, these results indicate that fluid intelligence is a primary driver of strategic level-k thinking in the beauty contest game. At the same time, as we discuss in Section 3.3, our reduced-form results provide evidence that the lower inclination of high judgment subjects to choose zero is driven by facets of judgment that are distinct from fluid intelligence.

[12]In specification (1), the effect of judgment is statistically significant at the 5% level using a one-sided test. In light of previous evidence that cognitive ability increases level-k (e.g., Gill and Prowse, 2016, Alós-Ferrer and Buckenmaier, 2021) and Kocher and Sutter (2005)'s conjecture that success in the beauty contest also requires judgment, it is plausible to formulate a one-tailed hypothesis that judgment enhances level-k.

[13]The effect size of cognitive ability here is close to that found by Gill and Prowse (2016).

	(1)	(2)	(3)
Panel I: Level-k type probabilities			
Level-0	0.064	0.080**	0.076**
	(0.042)	(0.033)	(0.036)
Level-1	0.434***	0.422***	0.424***
	(0.065)	(0.063)	(0.064)
Level-2	0.330***	0.327***	0.328***
	(0.057)	(0.059)	(0.057)
Level-3	0.172***	0.171***	0.172***
	(0.052)	(0.053)	(0.051)
Average level-k	1.611***	1.590***	1.596***
	(0.112)	(0.105)	(0.104)
Panel II: Marginal effects on average level-k			
Effect of judgment (1 SD increase)	0.147*		0.061
	(0.087)		(0.087)
Effect of cognitive ability (1 SD increase)		0.289***	0.266***
		(0.068)	(0.074)

Notes: We estimated the structural level-k model using the same maximum likelihood estimation routine as in Gill and Prowse (2016), and using all 999 subject-round observations. Across specifications, we varied the contents of $\mathbf{z}_i$: in specification (1) we included the subject's score on the judgment test, in specification (2) we included their score on the cognitive ability test, and in specification (3) we included scores on both tests. Each level-k type probability in Panel I was obtained by using the parameter estimates to evaluate the probability of that level-k type for each subject, conditional on the subject's observed $\mathbf{z}_i$ (see equation (1) above), and then averaging across subjects. Following Gill and Prowse (2016, Table 11), the average level-k in Panel I was obtained by computing the average level-k type for each subject, conditional on the subject's observed $\mathbf{z}_i$, and then averaging across subjects (to compute the average level-k type for a subject, we use the level-k type probabilities from equation (1)). In Panel II, marginal effects are averages of subject-level marginal effects, where each subject-level marginal effect was obtained by: (i) computing the average level-k type for the subject, conditional on the subject's observed $\mathbf{z}_i$; (ii) computing the average level-k type for the subject, after increasing their relevant test score in $\mathbf{z}_i$ by one standard deviation (1 SD); and (iii) then taking the difference. Standard errors were obtained from a Hessian matrix computed using numerical differentiation. ***, **, and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table 2: Structural level-k estimates

4 Further analyses

In this section, we present further analyses to understand more about how cognitive ability and judgment influence strategic behavior.

4.1 Choosing excessively low numbers

Table 3 provides evidence that subjects who are higher in cognitive ability are more likely to choose excessively low numbers. Specifically, in the second half, a one-standard-deviation increase in cognitive ability predicts a three-percentage-point increase in the probability of choosing below the best response (or lowest best response with multiplicity) to the subject's opponents' choices in the same round (two-sided $p = 0.020$ in specification (3) in Panel III). The effect size is large when compared to the average proportion of choices below the best response of 0.12 in the second half. This result suggests that some high cognitive ability subjects form beliefs about others' choices that are too low, although when interpreting this result we caution that we did not measure beliefs directly. To the extent that some high cognitive ability subjects form beliefs that are too low, these beliefs could be be skewed downward partly by an understanding that the Nash equilibrium action is zero, which would represent a type of "curse of knowledge" or "inability of people who have learned new information to imagine what not knowing the information is like" (Camerer et al., 2002).

By contrast, subjects who are higher in judgment are *less likely* to choose excessively low numbers. Specifically, in the second half, a one-standard-deviation increase in judgment predicts a three-percentage-point decrease in the probability of choosing below the best response to the subject's opponents' choices in the same round (two-sided $p = 0.030$ in specification (3) in Panel III of Table 3).

This tendency of higher cognitive ability subjects to choose excessively low numbers, which is reversed for higher judgment subjects, might provide a partial explanation for our results in Section 3.3 that subjects who are higher in cognitive ability are more likely to choose zero in the second half, while subjects who are higher in judgment are less likely to choose zero (see Table 1).

	Chose below best response to opponents' choices in current round		
	(1)	(2)	(3)
Panel I: All rounds			
Effect of judgment (1 SD increase)	-0.007		-0.014
	(0.009)		(0.009)
Effect of cognitive ability (1 SD increase)		0.017	0.021*
		(0.010)	(0.011)
Mean of dependent variable	0.11	0.11	0.11
Subject-round observations	999	999	999
Panel II: First half			
Effect of judgment (1 SD increase)	0.005		0.001
	(0.012)		(0.013)
Effect of cognitive ability (1 SD increase)		0.013	0.012
		(0.017)	(0.019)
Mean of dependent variable	0.10	0.10	0.10
Subject-round observations	564	564	564
Panel III: Second half			
Effect of judgment (1 SD increase)	-0.025**		-0.034**
	(0.012)		(0.014)
Effect of cognitive ability (1 SD increase)		0.022	0.032**
		(0.013)	(0.014)
Mean of dependent variable	0.12	0.12	0.12
Subject-round observations	435	435	435

Notes: See the notes to Table 1. Here, we run the same regressions, replacing the dependent variable with an indicator for whether the subject's choice in round r was below the lowest choice in the set of best responses to the choices of the subject's two opponents in that same round r.

Table 3: Effects of cognitive skills on proportion of choices below best response to opponents' choices in current round

4.2 Rule learners

In order to gain deeper insight into the relationship between cognitive skills and level-k thinking, we augment the structural level-k model to include rule learners who increase their level-k across repetitions of the beauty contest game. The structural model from Section 3.4 includes four level-k types (for $k \in \{0, 1, 2, 3\}$) who use the same level-k choice rule in every round $r \geq 2$ (recall that this model incorporates learning dynamics via the level-0 learning rule, with round 1 choices seeding the model). Here, we further include rule-learner types who switch from following the level-k choice rule in round 2 to following the level-$(k+1)$ choice rule in round 10.[14] Specifically, we further include three rule-learner types: L0-1, L1-2, and L2-3. Thus, the probability in equation (1) is now defined over seven types instead of four.

When we estimate this augmented structural level-k model, we find that a one-standard-deviation increase in judgment reduces the probability of being a rule learner by around thirteen percentage points (two-sided $p = 0.036$), while cognitive ability has a small and statistically insignificant effect on the probability of being a rule learner (two-sided $p = 0.653$).[15] When interpreting these results, we caution that we are pushing the limits of what we can estimate given our sample size. Indeed, in Section V.C, Gill and Prowse (2016) did not include rule-learner types when using their equation (6) (equation (1) here) to compare the effects of cognitive ability to those of personality.

In the level-k typology, a level-k type acts as if they believe that all others are of level-$(k-1)$. However, instead of being driven by explicit beliefs about the limited strategic sophistication of others, level-k behavior could instead arise from a cognitive bound that prevents the agent herself from reasoning at a higher level (e.g., Levin and Zhang, 2022). In this context, rule learning can arise in two potentially interrelated ways. First, a rule learner who is unconstrained by their cognitive bound might increase their belief

[14] We use the model of rule learning described in Section IV.A of Gill and Prowse (2016): for a Lk-$(k+1)$ rule leaner, the probability of choosing the level-$(k+1)$ choice rule increases linearly over rounds from 0 in round 2 to 1 in round 10, with the complementary probability of choosing the level-k choice rule falling linearly over the rounds. Footnote 17 in Gill and Prowse (2016) discusses this model of rule learning in more detail.

[15] We estimated the augmented model using the same methodology as described in the notes to Table 2. We used specification (3) that includes scores on both the judgment test and the cognitive ability test. We calculated marginal effects as described in the notes to Table 2, replacing the average level-k type for each subject with the probability of being a rule learner (i.e., with the sum of the probabilities of being each of the three rule-learner types).

about the strategic sophistication of others. Second, a rule learner's cognitive bound might loosen as they learn over time to reason at a higher level.

Viewed in this light, the finding described above that higher judgment subjects are less likely to be rule learners suggests that higher judgment subjects are more consistent or stable in their reasoning processes, in the sense that their own cognitive bound and/or beliefs about others' strategic sophistication are less likely to change over rounds. This stability of reasoning might provide a partial explanation for our results in Section 3.3 that high judgment subjects' probability of choosing zero responds less strongly to their group's average choice in the previous round (see Figure 3).

4.3 Whether zero was a best response in previous round

Finally, we dig deeper into the reduced-form results from Section 3.3 by repeating the analyses from Table 1 and Figure 3 of the probability of choosing zero in a round, but now factoring in whether zero was a best response to the choices of the subject's opponents in the previous round.

Recall that Table 1 provides evidence that subjects who are higher in cognitive ability are more likely to choose zero, while subjects who are higher in judgment are less likely to choose zero. Table 4 replicates Table 1, but now further including interactions between cognitive skills and an indicator for whether zero was a best response to the choices of the subject's two opponents in the previous round. Table 4 shows that the effects of cognitive ability and judgment are amplified when zero was a best response to the opponents' choices in the previous round. Specifically, focusing on the second half where effects are strongest, and including both cognitive skills in the same regression (specification (3) in Panel III), the positive effect of a one-standard-deviation increase in cognitive ability on the probability of choosing zero is sixteen percentage points stronger when zero was a best response to the opponents' choices in the previous round. Similarly, the negative effect of a one-standard-deviation increase in judgment on the probability of choosing zero is ten percentage points stronger when zero was a best response to the opponents' choices in the previous round. Panel I shows a similar pattern when considering all rounds.

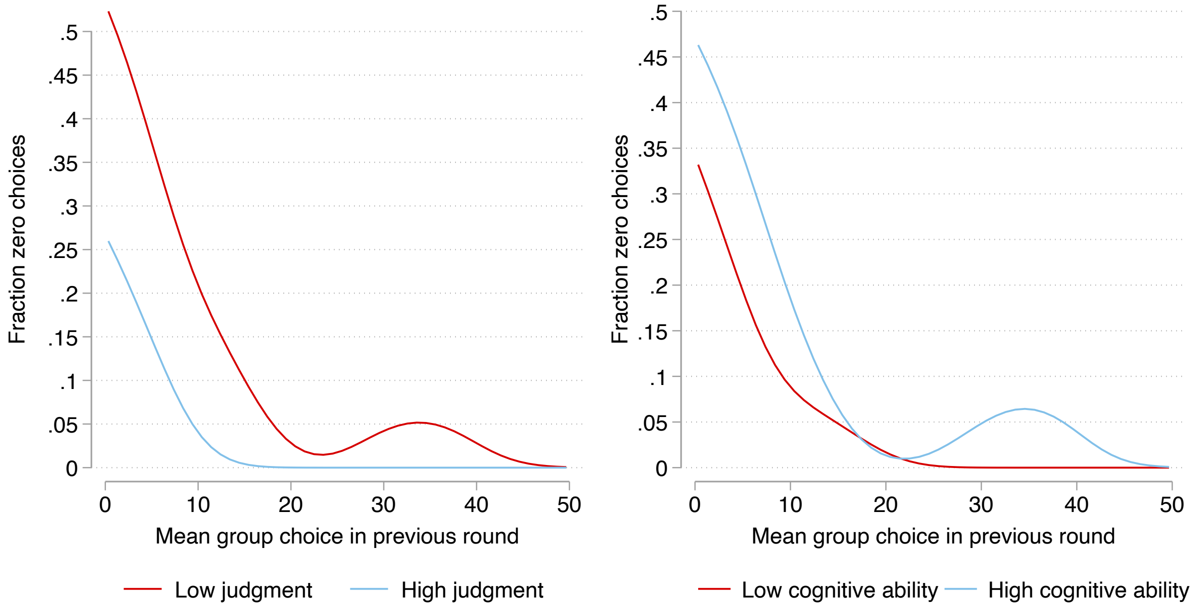
	Chose zero in round		
	(1)	(2)	(3)
Panel I: All rounds			
Judgment	-0.011**		-0.013**
	(0.005)		(0.006)
Judgment × Zero was a best response in previous round	-0.037		-0.080**
	(0.033)		(0.036)
Cognitive ability		0.006	0.010
		(0.006)	(0.007)
Cognitive ability × Zero was a best response in previous round		0.074**	0.106**
		(0.034)	(0.044)
Zero was a best response in previous round	0.151**	0.144**	0.142**
	(0.068)	(0.062)	(0.058)
Mean of dependent variable	0.05	0.05	0.05
Subject-round observations	858	858	858
Panel II: First half			
Judgment	-0.003		-0.005
	(0.003)		(0.004)
Judgment × Zero was a best response in previous round	-0.024		-0.030
	(0.024)		(0.030)
Cognitive ability		0.006	0.008
		(0.006)	(0.007)
Cognitive ability × Zero was a best response in previous round		0.001	0.014
		(0.004)	(0.015)
Zero was a best response in previous round	0.023	0.022	0.024
	(0.023)	(0.022)	(0.024)
Mean of dependent variable	0.01	0.01	0.01
Subject-round observations	423	423	423
Panel III: Second half			
Judgment	-0.018*		-0.022*
	(0.010)		(0.011)
Judgment × Zero was a best response in previous round	-0.041		-0.103**
	(0.049)		(0.044)
Cognitive ability		0.006	0.013
		(0.010)	(0.010)
Cognitive ability × Zero was a best response in previous round		0.118**	0.157**
		(0.044)	(0.059)
Zero was a best response in previous round	0.246**	0.227**	0.216***
	(0.101)	(0.085)	(0.079)
Mean of dependent variable	0.09	0.09	0.09
Subject-round observations	435	435	435

Notes: See the notes to Table 1. Here, we further include interactions between test scores and an indicator for whether zero was in the set of best responses to the choices of the subject's two opponents in the previous round (as well as the indicator itself). As in Table 1, the coefficients on test scores give the effect of a one-standard-deviation increase. The number of observations in Panels I and II fall compared to Table 1 because we can no longer include the subject's own choice in the first round.

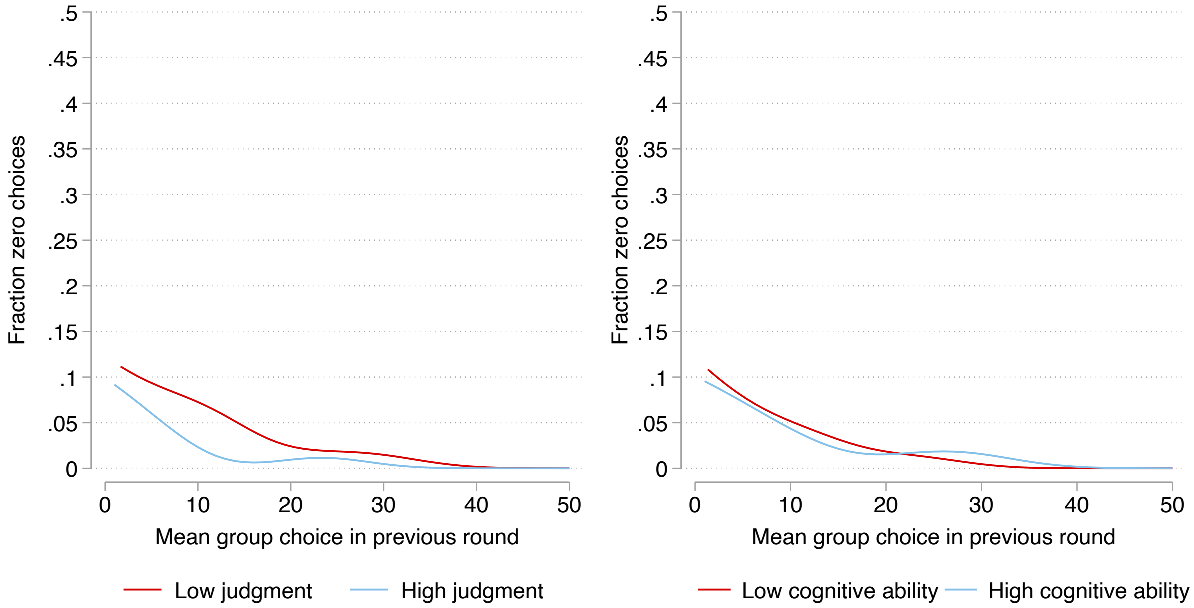
Table 4: Effects of cognitive skills on proportion of choices of zero:
By whether zero was a best response to opponents' choices in previous round

Recall that in Section 3.3 we uncover a dynamic mechanism that helps to shed light on why higher judgment subjects are less likely to choose zero. In particular, the left panel of Figure 3 illustrates that high judgment subjects' probability of choosing zero responds less strongly to their group's average choice in the previous round, compared to low judgment subjects. Figure 4 replicates Figure 3, but now splitting the data according to whether zero was a best response to the choices of the subject's opponents in the previous round. Consistent with the results from Table 4, when comparing the left panels of Figure 4(a) and Figure 4(b) we can see that the effects of judgment are concentrated in rounds in which zero was a best response to the opponents' choices in the previous round.

We concluded in Section 3.3 that the lower inclination of high judgment subjects to choose zero is driven by facets of judgment that are distinct from fluid intelligence. The analysis here in this subsection suggests that these facets of judgment play a particularly important role when zero was a best response to the choices of the subject's opponents in the previous round. By eliciting beliefs, a promising line of future research could measure the role of belief dynamics in explaining choices of zero and their dependence on whether zero was a best response in the previous round, and also study how these belief dynamics interact with different aspects of cognitive skills.



(a) Zero was a best response in previous round



(b) Zero was not a best response in previous round

Notes: See the notes to Figure 3, which shows the relationship between the proportion of choices of zero in round r and the mean choice in the subject's group of three in round $r-1$. Here, we split the data by whether zero was in the set of best responses to the choices of the subject's two opponents in round $r-1$.

Figure 4: Effect of mean group choice in previous round on proportion of choices of zero: By whether zero was a best response to opponents' choices in previous round

5 Conclusion

This study presents novel evidence about the roles that cognitive ability and judgment play in explaining strategic behavior. We present evidence that both cognitive ability and judgment help to predict behavior in the beauty contest game. Our findings indicate that fluid intelligence is a primary driver of strategic level-k thinking, while facets of judgment that are distinct from fluid intelligence drive the lower inclination of high judgment individuals to choose zero. Our results open up several interesting avenues for further research.

First, we find evidence of an intriguing dynamic mechanism, whereby judgment predicts how strongly subjects' probability of choosing zero responds to the group's average choice in the previous round. Designing a follow-up experiment to understand more about this mechanism would be valuable. For example, a new experiment could discover which facets of judgment are most at play and whether other cognitive skills also link to this mechanism.

Second, the roles of other higher-order cognitive skills in shaping strategic behavior have not been fully explored in the literature. For example, studying the influence of problem-solving, decision-making ability and creativity, in conjunction with cognitive ability and judgment, could provide a more comprehensive understanding of the factors that influence strategic behavior.

Third, future research could apply our findings to real-world strategic scenarios. The beauty contest game represents a simplified version of the complex decision-making situations individuals face in their everyday interactions. Future research could focus on the roles of cognitive ability and judgment in explaining financial decision-making, negotiation, strategies for building long-term relationships, and competitive behavior in the labor market.

References

Alaoui, L. and **Penta, A.** (2016). Endogenous depth of reasoning. *Review of Economic Studies*, 83(4): 1297–1333

Alaoui, L. and **Penta, A.** (2022). Cost-benefit analysis in reasoning. *Journal of Political Economy*, 130(4): 881–925

Alós-Ferrer, C. and **Buckenmaier, J.** (2021). Cognitive sophistication and deliberation times. *Experimental Economics*, 24(2): 558–592

Autor, D.H. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, 29(3): 3–30

Ballester, C., **Rodríguez-Moral, A.**, and **Vorsatz, M.** (2023). Cognitive reflection in experimental anchored guessing games. *SSRN Working Paper 4479869*

Ballester, C., **Vorsatz, M.**, and **Ponti, G.** (2021). Uncovering seeds. *MPRA Working Paper 117235*

Brañas-Garza, P., **García-Muñoz, T.**, and **González, R.H.** (2012). Cognitive effort in the beauty contest game. *Journal of Economic Behavior & Organization*, 83(2): 254–260

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

Burnham, T.C., **Cesarini, D.**, **Johannesson, M.**, **Lichtenstein, P.**, and **Wallace, B.** (2009). Higher cognitive ability is associated with lower entries in a p-beauty contest. *Journal of Economic Behavior & Organization*, 72(1): 171–175

Camerer, C.F., **Ho, T.H.**, and **Chong, J.K.** (2002). Sophisticated experience-weighted attraction learning and strategic teaching in repeated games. *Journal of Economic Theory*, 104(1): 137–188

Capra, C.M. (2019). Understanding decision processes in guessing games: a protocol analysis approach. *Journal of the Economic Science Association*, 5(1): 123–135

Carpenter, J., **Graham, M.**, and **Wolf, J.** (2013). Cognitive ability and strategic sophistication. *Games and Economic Behavior*, 80: 115–130

Carpenter, P.A., **Just, M.A.**, and **Shell, P.** (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices test. *Psychological Review*, 97(3): 404–413

Chan, D. and **Schmitt, N.** (2002). Situational judgment and job performance. *Human Performance*, 15(3): 233–254

Charness, G., **Rustichini, A.**, and **Van de Ven, J.** (2018). Self-confidence and strategic behavior. *Experimental Economics*, 21(1): 72–98

Chen, D.L., **Schonger, M.**, and **Wickens, C.** (2016). oTree–An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9: 88–97

Clevenger, J., **Pereira, G.M.**, **Wiechmann, D.**, **Schmitt, N.**, and **Harvey, V.S.** (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology*, 86(3): 410–417

Crawford, V.P., **Costa-Gomes, M.A.**, and **Iriberri, N.** (2013). Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. *Journal of Economic Literature*, 51(1): 5–62

Deming, D.J. (2021). The growing importance of decision-making on the job. *NBER Working Paper 28733*

Deming, D.J. (2022). Four facts about human capital. *Journal of Economic Perspectives*, 36(3): 75–102

Duffy, J. and **Nagel, R.** (1997). On the robustness of behaviour in experimental 'Beauty Contest' games. *Economic Journal*, 107(445): 1684–1700

Dworak, E.M., **Revelle, W.**, **Doebler, P.**, and **Condon, D.M.** (2021). Using the International Cognitive Ability Resource as an open source tool to explore individual differences in cognitive ability. *Personality and Individual Differences*, 169: 109906

Fe, E., **Gill, D.**, and **Prowse, V.** (2022). Cognitive skills, strategic sophistication, and life outcomes. *Journal of Political Economy*, 130(10): 2643–2704

Fehr, D. and **Huck, S.** (2016). Who knows it is a game? On strategic awareness and cognitive ability. *Experimental Economics*, 19(4): 713–726

Gill, D. and **Prowse, V.** (2016). Cognitive ability, character skills and learning to play equilibrium: A level-k analysis. *Journal of Political Economy*, 124(6): 1619–1676

Gill, D. and **Prowse, V.** (2023). Strategic complexity and the value of thinking. *Economic Journal*, 133(650): 761–786

Gill, D. and **Prowse, V.** (forthcoming). The creativity premium: Exploring the link between childhood creativity and life outcomes. *Journal of Political Economy Microeconomics*

Gill, D. and **Rosokha, Y.** (forthcoming). Beliefs, learning, and personality in the indefinitely repeated prisoner's dilemma. *American Economic Journal: Microeconomics*

Gray, J. and **Thompson, P.** (2004). Neurobiology of intelligence: Science and ethics. *Nature Reviews Neuroscience*, 5(6): 471–482

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1): 114–125

Hermes, H. and **Schunk, D.** (2022). If you could read my mind – an experimental beauty-contest game with children. *Experimental Economics*, 25(1): 229–253

Ho, T.H., **Camerer, C.**, and **Weigelt, K.** (1998). Iterated dominance and iterated best response in experimental "*p*-beauty contests". *American Economic Review*, 88(4): 947–969

Kocher, M.G. and **Sutter, M.** (2005). The decision maker matters: Individual versus group behaviour in experimental beauty-contest games. *Economic Journal*, 115(500): 200–223

Levin, D. and **Zhang, L.** (2022). Bridging level-k to Nash equilibrium. *Review of Economics and Statistics*, 104(6): 1329–1340

Lievens, F., **Peeters, H.**, and **Schollaert, E.** (2008). Situational judgment tests: A review of recent research. *Personnel Review*, 37(4): 426–441

McDaniel, M.A., **Hartman, N.S.**, **Whetzel, D.L.**, and **Grubb III, W.L.** (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, 60(1): 63–91

McDaniel, M.A. and **Nguyen, N.T.** (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment*, 9(1-2): 103–113

McDaniel, M.A., **Psotka, J.**, **Legree, P.J.**, **Yost, A.P.**, and **Weekley, J.A.** (2011). Toward an understanding of situational judgment item validity and group differences. *Journal of Applied Psychology*, 96(2): 327–336

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *American Economic Review*, 85(5): 1313–1326

Nagel, R., **Bühren, C.**, and **Frank, B.** (2017). Inspired and inspiring: Hervé Moulin and the discovery of the beauty contest game. *Mathematical Social Sciences*, 90: 191–207

Oostrom, J.K., **De Soete, B.**, and **Lievens, F.** (2015). Situational judgment testing: A review and some new developments. In I. Nikolaou and J.K. Oostrom, editors, *Employee Recruitment, Selection, and Assessment: Contemporary Issues for Theory and Practice*, 172–189. Psychology Press

Proto, E., **Rustichini, A.**, and **Sofianos, A.** (2019). Intelligence, personality and gains from cooperation in repeated interactions. *Journal of Political Economy*, 127(3): 1351–1390

Proto, E., **Rustichini, A.**, and **Sofianos, A.** (2022). Intelligence, errors, and cooperation in repeated interactions. *Review of Economic Studies*, 89(5): 2723–2767

Raven, J., **Raven, J.C.**, and **Court, J.H.** (2000). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* San Antonio: Pearson

Rubinstein, A. (2016). A typology of players: Between instinctive and contemplative. *Quarterly Journal of Economics*, 131(2): 859–890

Schmitt, N. and **Chan, D.** (2006). Situational judgment tests: Method or construct? In J. Weekley and R. Ployhart, editors, *Situational Judgment Tests: Theory, Measurement, and Application*, 135–156. Lawrence Erlbaum Associates

Schnusenberg, O. and **Gallo, A.** (2011). On cognitive ability and learning in a beauty contest. *Journal for Economic Educators*, 11(1): 13–24

Stahl, D.O. (1996). Boundedly rational rule learning in a guessing game. *Games and Economic Behavior*, 16(2): 303–330

Stahl, D.O. and **Wilson, P.W.** (1994). Experimental evidence on players' models of other players. *Journal of Economic Behavior & Organization*, 25(3): 309–327

Stemler, S.E. and **Sternberg, R.J.** (2006). Using situational judgment tests to measure practical intelligence. In J. Weekley and R. Ployhart, editors, *Situational Judgment Tests: Theory, Measurement, and Application*, 107–131. Lawrence Erlbaum Associates

Sternberg, R.J. (1999). The theory of successful intelligence. *Review of General Psychology*, 3(4): 292–316

Sternberg, R.J., **Wagner, R.K.**, **Williams, W.M.**, and **Horvath, J.A.** (1995). Testing common sense. *American Psychologist*, 50(11): 912

Taylor, M.P. (2020). Heterogeneous motivation and cognitive ability in the lab. *Journal of Behavioral and Experimental Economics*, 85: 101523

Weekley, J.A. and **Jones, C.** (1997). Video-based situational testing. *Personnel Psychology*, 50(1): 25–49

Weekley, J.A. and **Jones, C.** (1999). Further studies of situational tests. *Personnel Psychology*, 52(3): 679–700

Weekley, J.A. and **Ployhart, R.E.** (2005). Situational judgment: Antecedents and relationships with performance. *Human Performance*, 18(1): 81–104

Whetzel, D.L. and **McDaniel, M.A.** (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review*, 19(3): 188–202

Web Appendix

(Intended for Online Publication)

Web Appendix I Further information on experiment

Web Appendix I.1 Experimental instructions

[Screen 1] Introduction. Please now turn off cell phones and any other electronic devices. These must remain turned off for the duration of this session. Please do not use or place on your desk any personal items, including pens, paper, phones etc. Please do not look into anyone else's booth at any time. Thank you for participating in this experimental session on economic decision-making. You were randomly selected from the Vernon Smith Experimental Economics Laboratory's pool of subjects to be invited to participate in this session. There will be a number of points in these instructions where you will be asked to raise your hand if you have any questions. Apart from asking questions in this way, you must not communicate with anybody in this room or make any noise. You will be paid a show-up fee of $5 together with any money you accumulate during this session. The amount of money you accumulate will depend partly on your actions and partly on the actions of other participants. You will be paid privately in cash at the end of the session. Please raise your hand if you have any questions. [Button: Click to continue.]

[Screen 2] Introduction. This session is made up of 2 parts. In the first part of the session, you will participate in an economic interaction that includes opportunities to earn money. In the second part of the session, you will be asked to complete two tests and a short questionnaire. We will pay you $3 for each test (irrespective of your scores on the tests). You will not be paid for completing the questionnaire. Please raise your hand if you have any questions. [Button: Click to continue.]

[Screen 3] Instructions on Part 1 of the session. We now describe the economic interaction that makes up the first part of the session. The economic interaction will last for up to 10 rounds. You will be anonymously matched into groups of 3 participants. You will stay in the same group for all rounds. In each round, you and your other 2 group members will separately choose a whole number between 0 and 100 (0, 100 or any whole number in between is allowed). The group member whose chosen number is closest to 70% of the average of all 3 chosen numbers will be paid $6 for that round and the other 2 group members will be paid nothing. If more than one group member chooses a number which is closest to 70% of the average of all 3 chosen numbers, the $6 will be split equally among the group members who chose the closest number or numbers. Your total payment from the economic interaction will be the sum of your payments in each round. In each round you will have 90 seconds to choose your number. If you choose your number early you will still have to wait until the end of the 90 seconds. The screen will display the time remaining (in seconds). The screen will also include a reminder of the rules. At the end of each round you will discover: (i) the numbers chosen by all your group members; (ii) the average of all 3 chosen numbers; (iii) what 70% of the average of all 3 chosen numbers was; (iv) how much each group member will be paid for the round. Please raise your hand if you have any questions. [Button: Click to continue.]

[Screen 4] Instructions on Part 1 of the session. Recall, the economic interaction will last for up to 10 rounds. Specifically: (1) The economic interaction will last for at least 5 rounds. (2) Starting from the 5th round, at the end of each round there is a [depending on treatment: 50%, 75%, or 90%] chance that the economic interaction continues to the next round and a [depending on treatment: 50%, 25%, or 10%] chance that the economic interaction ends. (3) If the economic interaction reaches the 10th round, then the interaction ends for sure at the end of that 10th round. You will stay in the same group of 3 for all rounds of the economic interaction. Each group member has been randomly allocated a label, X, Y or Z. Your label is shown below. [Example: You are group member X.] Please raise your hand if you have any questions. There will be no further opportunities for questions on this part of the session. [Button: Click to continue.]

[Between 5 and 10 rounds of the p-beauty contest game.]

[Subjects complete two tests of cognitive skill and a demographic questionnaire.]

[Final screen] The session has now finished. Your total cash payment, including the show-up fee, is [total payment in dollars]. Please remain seated in your booth. The lab assistant will come to your booth to give you your cash payment. Thank-you for participating.

Web Appendix I.2 Further details about p-beauty contest game

Each group of three subjects played the p-beauty contest game between five and ten times: (i) each group played five rounds for sure; (ii) starting from round 5, the game continued to the next round with probability $q < 1$; and (iii) if the game reached round 10, the game ended for sure at the end of that tenth round. We ran twelve sessions, with an average of

11.75 subjects per session, and with four sessions for each value of $q \in \{0.5, 0.75, 0.9\}$. We included variation in q because Online Appendix II.6 of Gill and Prowse (2023) also uses the same dataset to show that subjects are not forward-looking in the p-beauty contest; however, the cognitive ability and judgment measures that are central to this paper are not analyzed there.

Web Appendix I.3 Further details about tests and questionnaire

The judgment test lasted 20 minutes. We paid subjects $3 for completing the test, but following the convention in the psychometric literature we did not pay subjects for their performance in the test. We used eleven items from the seventy-two situational judgment items from ICAR. In particular, we selected items from the twenty that measure situational judgment only in the domain of "analysis and problem solving" because that domain is most comparable to cognitive ability as measured by the matrix reasoning test. Among those twenty items, we excluded five whose theme is "secretarial" or "organizational" instead of "administrative". Among the remaining fifteen items, we excluded four that use either acronyms that may be unfamiliar to our subjects or gendered first names. Specifically, we used the following eleven items: 1; 8; 16; 17; 24; 30; 31; 34; 46; 54; 59. Each item has one best answer that scores one point, one worst answer that scores zero points, and two answers that score 0.5 points. Web Appendix I.4 provides one of the items as a sample and Web Appendix I.5 provides a brief description of the eleven situations used in our eleven-item test of situational judgment.

The cognitive ability test lasted 10 minutes. We paid subjects $3 for completing the test, but following the convention in the psychometric literature we did not pay subjects for their performance in the test. We used the full eleven-item matrix reasoning test from ICAR (Gill and Rosokha, forthcoming, also use the test in a setting with strategic behavior). Web Appendix I.6 provides one of the items as a sample.

The demographic questionnaire asked about gender, age (18-25; 26-40; 41-60; 61+), whether English is the subject's native language, and if not, age subject started learning English (0-5; 6-10; 11-15; 16+).

Web Appendix I.4 Sample item from judgment test

Your department is currently undergoing an IT modernisation programme. Although you are not a technical expert yourself, you are given the responsibility to introduce the new IT system to your department. You have already spent almost three weeks learning the specifics and giving introductory sessions to staff members, when the technical team identifies bugs in the new IT system and decides to cancel the system replacement. Not only was your work unnecessary, but the staff members also appear frustrated with you now.

In your opinion, which of the following is the best way to manage this situation:

1. You speak with the technical team and try to resolve the issue and continue with the system replacement. Your organisation has invested too much time and effort already for this change to cancel the replacement now.

2. You go home and start your own research in order to remove the bugs issue and maintain the system. Your organisation has invested too much time and effort already for this change to cancel the replacement now.

3. You speak with your boss and with the technical team to gauge whether going back to the old system is the best choice. After that, you communicate the new plan to your staff members and discuss with them any potential concerns they may have.

4. You speak with the staff members and use your persuasion skills to convince them that every choice made by the organisation is for the best, even if sometimes it does not end up being the most efficient one.

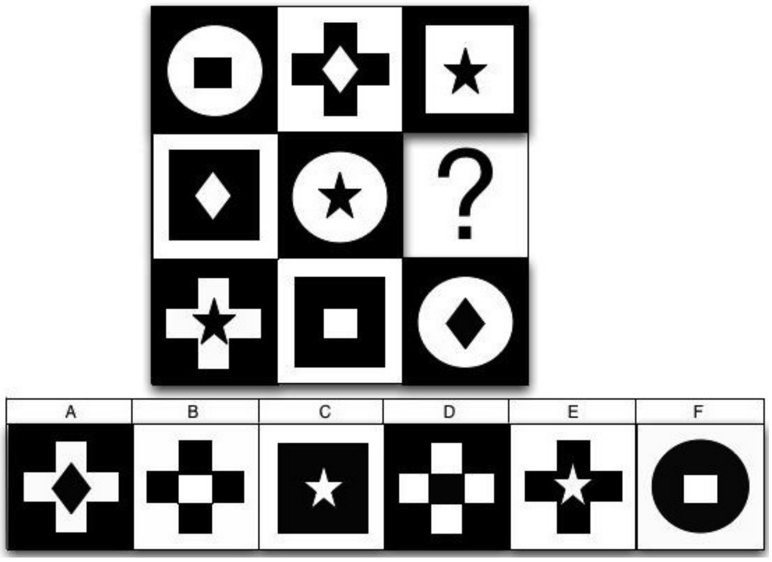
Web Appendix I.5 Description of situations from judgment test

Below we provide a brief description of the eleven situations used in the eleven-item test of situational judgment:

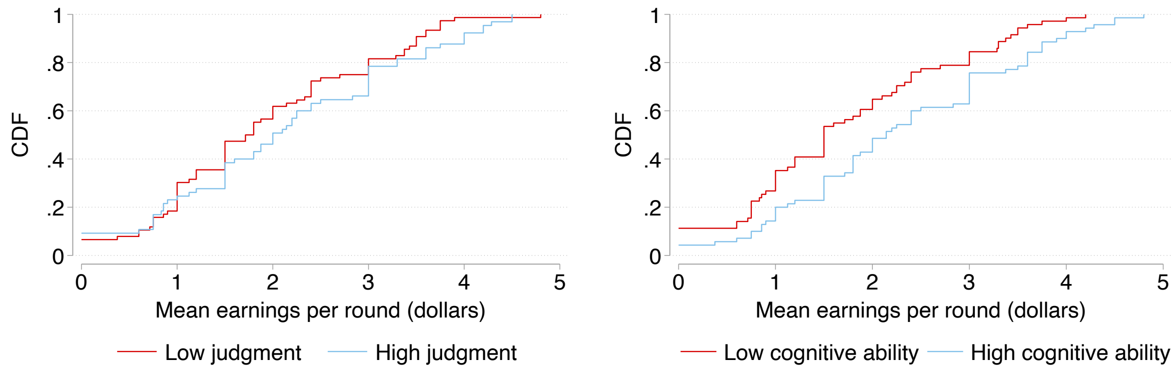
- Managing a team during a period of forced redundancies.
- Managing skilled workers who are known to have difficulties keeping to deadlines.
- Working with an overly critical superior.
- Working with a superior who places blame perceived to be unfair.
- Dealing with the problematic introduction of a new workplace technology system.
- Managing a team whose work was sent back for revision by a superior due to an important oversight.
- Managing a team whose relocation has led to disorganization and confusion.
- Dealing with the discovery of a fundamental error a couple of days before the deadline to present the results of a project in person.
- Dealing with the last-minute unavailability of the planned presenter of a group project.
- Managing the collaboration between two teams that lack trust in one another.
- Managing a team whose morale has fallen due to a relocation to a less favorable work location.

Web Appendix I.6 Sample item from cognitive ability test

Please indicate which is the best answer to complete the figure below. There is only one right answer.

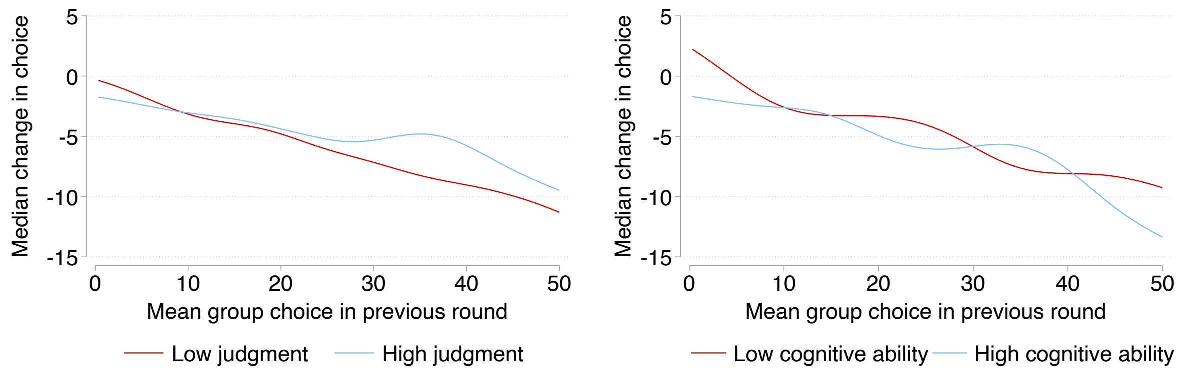


Web Appendix II   Further figures and tables



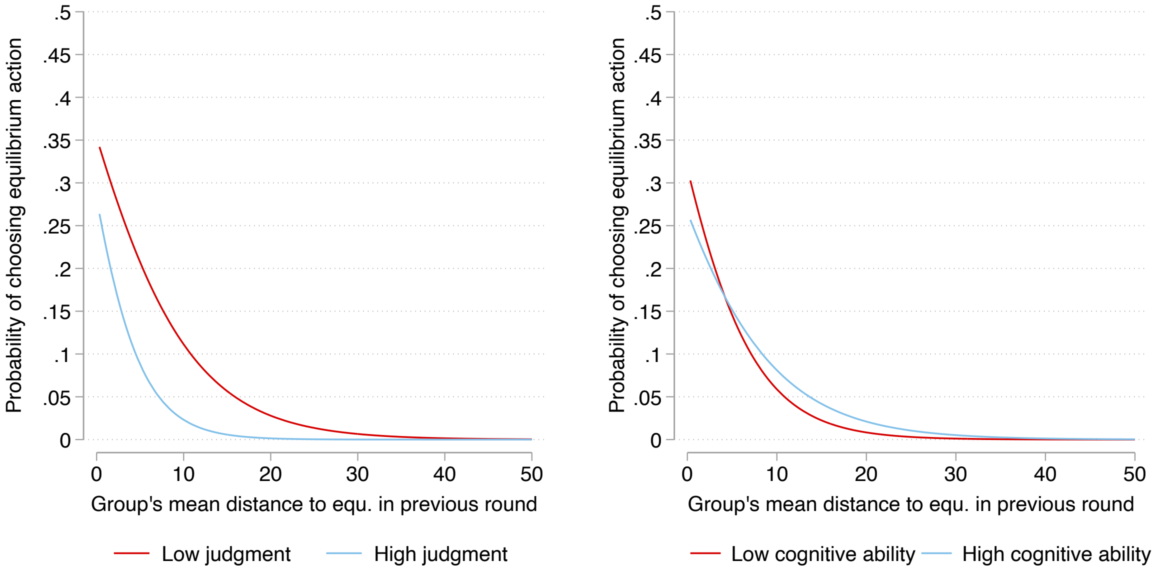
Notes: The notes to Figure 2 describe how we classify subjects by cognitive skill level.

Figure A.1: CDFs of subject-level mean earnings across rounds



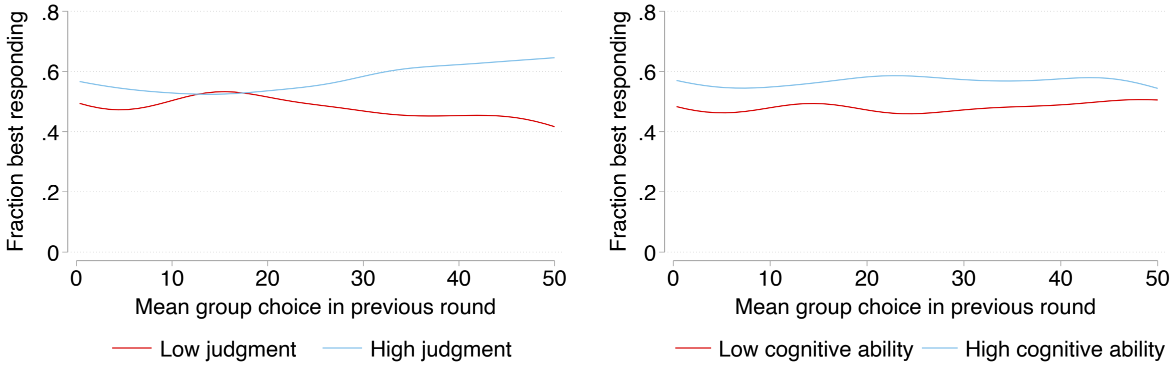
Notes: See the notes to Figure 3, which shows the relationship between the proportion of choices of zero in round r and the mean choice in the subject's group of three in round $r-1$. Here, on the vertical axis we replace the proportion of choices of zero in round r with the median change in choice between round $r-1$ and round r.

Figure A.2: Effect of mean group choice in previous round on median change in choice

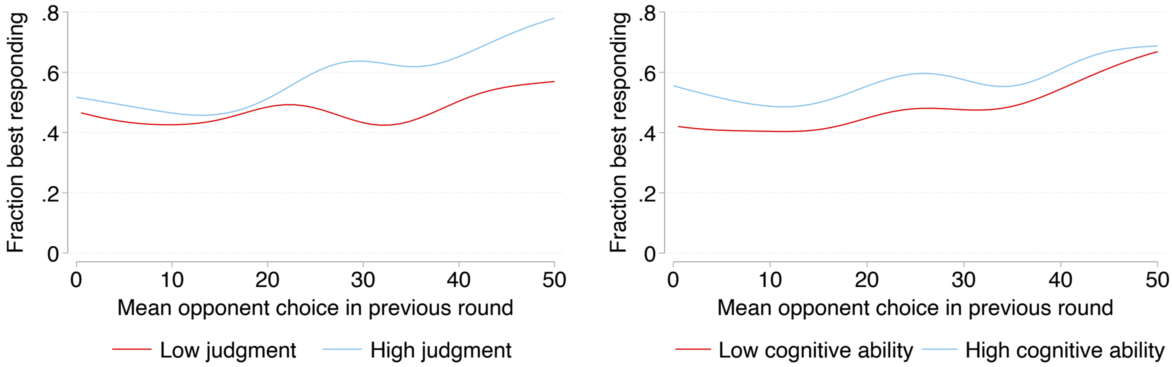


Notes: See the notes to Figure 3, which shows the relationship between the proportion of choices of zero in round r and the mean choice in the subject's group of three in round $r-1$. Here, we make two changes. First, we interpret this relationship in the context of equilibrium, recalling that zero is the equilibrium action, and noting therefore that the mean choice in the subject's group in round $r-1$ (on the horizontal axis in Figure 3) equals the mean distance to the equilibrium action in the subject's group in round $r-1$ (on the horizontal axis here). Second, we add more structure by replacing the proportion of choices of zero in round r (on the vertical axis in Figure 3) with estimates of the probability of choosing the equilibrium action in round r (on the vertical axis here) that come from a binary logistic probability model estimated using maximum likelihood. Specifically, letting e (n) represent choosing (not choosing) the equilibrium action, $\Pr(e|\mathbf{X}_{i,r}) = \frac{\exp(\boldsymbol{\beta}_e \mathbf{X}_{i,r})}{\sum_{j \in \{e,n\}} \exp(\boldsymbol{\beta}_j \mathbf{X}_{i,r})}$, where $\mathbf{X}_{i,r}$ is a vector that contains the mean distance to the equilibrium action in subject i's group in round $r-1$, the subject's score on the judgment test, their score on the cognitive ability test, the interaction of these scores and the mean distance to the equilibrium action in the subject's group in round $r-1$, and a constant (we impose the identifying normalization that $\boldsymbol{\beta}_n = \mathbf{0}$).

Figure A.3: Effect of group's mean distance to equilibrium in previous round on logistic probability of choosing the equilibrium action



(a) By mean group choice in previous round



(b) By mean of opponents' choices in previous round

Notes: See the notes to Figure 3, which shows the relationship between the proportion of choices of zero in round r and the mean choice in the subject's group of three in round $r-1$. Here, in subfigure (a), on the vertical axis we replace the proportion of choices of zero in round r with the proportion of choices in round r that were in the set of best responses to the choices of the subject's two opponents in round $r-1$. In subfigure (b), we retain the same vertical axis as in subfigure (a), while on the horizontal axis we replace the mean choice in the subject's group of three in round $r-1$ with the mean choice of the subject's two opponents in round $r-1$.

Figure A.4: Effect of group choices in previous round on proportion of current round choices that best respond to opponents' previous round choices

	Chose one or two in round		
	(1)	(2)	(3)
Panel I: All rounds			
Effect of judgment (1 SD increase)	0.008		0.003
	(0.010)		(0.011)
Effect of cognitive ability (1 SD increase)		0.014*	0.013
		(0.008)	(0.009)
Mean of dependent variable	0.08	0.08	0.08
Subject-round observations	999	999	999
Panel II: First half			
Effect of judgment (1 SD increase)	0.006		0.002
	(0.005)		(0.006)
Effect of cognitive ability (1 SD increase)		0.012*	0.011
		(0.007)	(0.008)
Mean of dependent variable	0.02	0.02	0.02
Subject-round observations	564	564	564
Panel III: Second half			
Effect of judgment (1 SD increase)	0.010		0.005
	(0.024)		(0.026)
Effect of cognitive ability (1 SD increase)		0.017	0.015
		(0.019)	(0.021)
Mean of dependent variable	0.15	0.15	0.15
Subject-round observations	435	435	435

Notes: See the notes to Table 1. Here, we run the same regressions, but replacing the dependent variable with an indicator for whether the subject chose one or two in the round.

Table A.1: Effects of cognitive skills on proportion of choices of one or two

Web Appendix III Evidence on intuitive thinking

An insightful referee asked about the role of intuition, which they "characterized as a faster, more unconscious process." Specifically, this referee: (i) noted that there appears to be a lack of formal models that capture the role of intuitive thinking in decision-making; (ii) remarked that intuition could potentially be a significant factor in level-k strategic thinking; and (iii) asked whether choices of zero in our setting can be explained by intuitive decision-making.

Response time (also known as decision time) is an indicator of intuitive or instinctive thinking (e.g., Rubinstein, 2016). As described by Gill and Prowse (2023): "Response times are thought to be connected to decision-making style: fast thinking is linked to intuitive or instinctive decision-making, while slower thinking is linked to a more deliberate or contemplative mode of thought..."

Using data on response times, the existing literature provides suggestive evidence that higher level-k thinking is associated with more deliberate, less intuitive decision-making. Specifically, Alós-Ferrer and Buckenmaier (2021) and Gill and Prowse (2023) find that higher level-k types think for longer (in the one-shot and repeated beauty contest game, respectively). A positive relationship between level-k and deliberate thinking is consistent with Alaoui and Penta (2016, 2022)'s model of endogenous depth of reasoning, in which the number of steps of reasoning is determined by a comparison of the cost and value of each additional step.

However, turning specifically to choices of zero, we present three pieces of evidence which suggest that these choices of zero are not connected to intuitive thinking. The good fit of the structural level-k model that we borrow from Gill and Prowse (2016) suggests that, instead, the learning dynamics in this level-k model play an important role in driving choices of zero.[16] Having said this, we did not design our experiment to study the role of intuition, and so the evidence below should be not be interpreted as definitive.

[16] Using a dataset of 780 subjects repeatedly playing the p-beauty contest game (with the same p and fixed groups of three that we use here), Gill and Prowse (2016) provide substantial evidence of goodness of fit, including measures of fit based on convergence of choices to zero.

The first piece of evidence which suggests that choices of zero are not connected to intuitive thinking comes from Table A.2, which adds response time to the regressions from Table 1 in Section 3.3 that studied choices of zero.[17] Specifically:

(a) The lack of statistical significance of the coefficients on response time in Column 1 of Table A.2 provides evidence that response times do not predict the likelihood of choosing zero.

(b) The fact that the coefficients on cognitive skills in columns 2 to 4 of Table A.2 are almost identical to those in columns 1 to 3 of Table 1 provides evidence that controlling for response times does not affect the relationship between cognitive skills and choices of zero.

(c) The lack of statistical significance of the interaction terms in column 5 of Table A.2 provides evidence that response times do not mediate the relationship between cognitive skills and choices of zero.

The second piece of evidence which suggests that choices of zero are not connected to intuitive thinking comes from Figure A.5 that shows the distribution of choices, split by first half versus second half. In particular, Figure A.5 shows that choices of zero were uncommon in the first half of the experiment, and so do not appear to be intuitive before learning takes place. This evidence also suggests that zero is not a focal point absent learning dynamics.

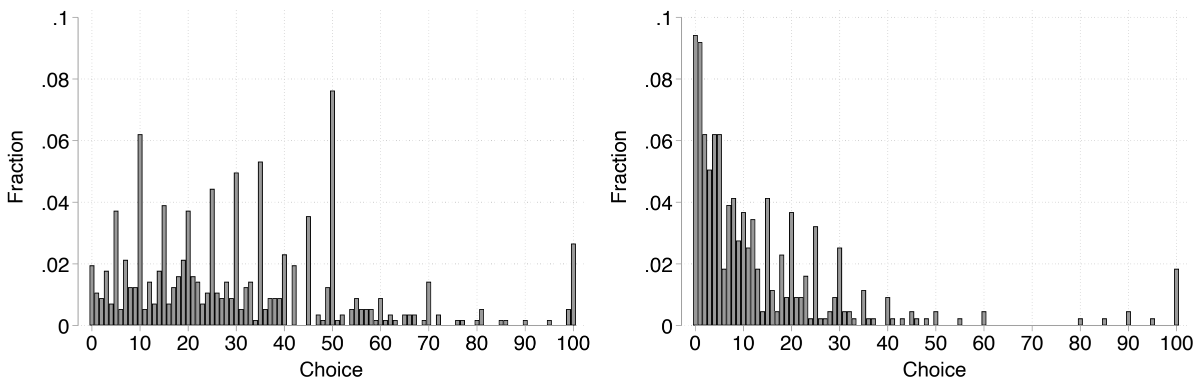
The third piece of evidence which suggests that choices of zero are not connected to intuitive thinking comes from Gill and Prowse (2023)'s analysis of response times (that uses the dataset from Gill and Prowse, 2016, described in footnote 16). In particular, Gill and Prowse (2023, Online Appendix VI.2) find that when a group of three gets close to zero, or plays zero for the first time, average response times are similar to the average for all choices in the dataset. By contrast, average response times fall substantially when the group played zero in the previous round (presumably because the equilibrium logic then becomes obvious).

[17] As stated in the experimental instructions in Web Appendix I.1, subjects in our experiment had 90 seconds to make their choice. This design feature mirrors Gill and Prowse (2023), who study response times in detail. Following Gill and Prowse (2023), in Table A.2 we include the effect of response time measured in minutes.

	Chose zero in round				
	(1)	(2)	(3)	(4)	(5)
Panel I: All rounds					
Response time	0.006	0.011	0.000	0.005	0.010
	(0.023)	(0.024)	(0.024)	(0.023)	(0.024)
Judgment		-0.016*		-0.024**	-0.024***
		(0.009)		(0.011)	(0.009)
Cognitive ability			0.019*	0.026*	0.037*
			(0.011)	(0.014)	(0.019)
Judgment × Response time					0.001
					(0.026)
Cognitive ability × Response time					-0.034
					(0.025)
Mean of dependent variable	0.05	0.05	0.05	0.05	0.05
Subject-round observations	999	999	999	999	999
Panel II: First half					
Response time	0.033	0.035	0.031	0.032	0.032
	(0.026)	(0.028)	(0.025)	(0.026)	(0.026)
Judgment		-0.006		-0.009	-0.013*
		(0.007)		(0.009)	(0.006)
Cognitive ability			0.005	0.008	0.009
			(0.006)	(0.008)	(0.011)
Judgment × Response time					0.011
					(0.024)
Cognitive ability × Response time					-0.004
					(0.014)
Mean of dependent variable	0.02	0.02	0.02	0.02	0.02
Subject-round observations	564	564	564	564	564
Panel III: Second half					
Response time	-0.025	-0.015	-0.032	-0.019	0.006
	(0.047)	(0.046)	(0.049)	(0.046)	(0.051)
Judgment		-0.030*		-0.045**	-0.032**
		(0.015)		(0.017)	(0.014)
Cognitive ability			0.039	0.052*	0.070**
			(0.025)	(0.027)	(0.034)
Judgment × Response time					-0.043
					(0.045)
Cognitive ability × Response time					-0.075
					(0.048)
Mean of dependent variable	0.09	0.09	0.09	0.09	0.09
Subject-round observations	435	435	435	435	435

Notes: See the notes to Table 1. The regressions in columns (2) to (4) here are the same as those in columns (1) to (3) of Table 1, except that we add response time (that is, time elapsed before the subject chooses, measured in minutes) as an independent variable. The regression in column (5) here further adds interactions between test scores and response time, while the regression in column (1) here omits the test scores. As in Table 1, the coefficients on test scores give the effect of a one-standard-deviation increase.

Table A.2: Effects of cognitive skills and response time on proportion of choices of zero



(a) First half

(b) Second half

Notes: First and second halves are defined in the notes to Table 1.

Figure A.5: Distribution of choices

Web Appendix IV Role of information and feedback

An insightful referee asked about the role of information and feedback. Specifically, this referee noted that: (i) our experimental setting with feedback puts an emphasis on judgment as the ability to interpret information about past choices and payoffs in the subject's group; while (ii) judgment could play a role independently of the ability to use feedback, e.g., via the ability to leverage external information.

This referee makes a good point. Even in our setting with repetition and feedback, effects of judgment and cognitive ability could arise partly from leveraging external information held by subjects, as opposed to learning from feedback during the repeated game (that is, learning from experiencing payoffs or observing decisions within the subject's group). We provide two pieces of evidence which suggest that the effects of cognitive skills that we observe are driven mainly by learning from feedback, although we do not rule out some independent role for leveraging external information, and our data are not definitive on this point.

The first piece of evidence which suggests that the effects of cognitive skills that we observe are driven mainly by learning from feedback comes from our reduced-form results. The analyses that split the results by rounds (specifically by first half versus second half) are presented in Table 1 in Section 3.3 and Table A.1 in Web Appendix II. As explained in the notes to Table 1, the first half includes rounds 1 to 4, and so includes only a maximum of three rounds of feedback. These tables show that in the first half of the experiment, when the amount of feedback is still limited, neither judgment nor cognitive ability has a statistically significant effect (on choices of zero in Panel II of Table 1 or on close-to-zero choices in Panel II of Table A.1). Instead, the effects of cognitive skills become apparent in the second half of the experiment (on choices of zero in Panel III of Table 1). Table A.3 and Table A.4 replicate theses analyses, but now splitting the results by round 1 versus later rounds (i.e., rounds 2 to 10), and show that cognitive skills do not have statistically significant effects in round 1 before subjects receive any feedback. However, these round 1 results should be interpreted with considerable caution, since we are underpowered to detect effects using small numbers of rounds.

The second piece of evidence which suggests that the effects of cognitive skills that we observe are driven mainly by learning from feedback comes from the good fit of the structural level-k model from Gill and Prowse (2016), which includes cognitive ability and that we borrow here. Using a dataset of 780 subjects repeatedly playing the p-beauty contest game (with the same p and fixed groups of three that we use here), Gill and Prowse (2016) provide substantial evidence of goodness of fit. This good fit suggests that the learning dynamics in this level-k model interact with cognitive ability to drive behavior, although for our purposes a limitation of this evidence is that Gill and Prowse (2016)'s model includes only cognitive ability (with judgment absent).

Finally, we clarify that we are not able to estimate the structural level-k model only for round 1: the reason is that round 1 choices seed the model by serving as the initial conditions for the learning dynamics. Furthermore, due to limitations imposed by the number of observations in our dataset, we are unable estimate the structural model only for round 2.

	Chose zero in round		
	(1)	(2)	(3)
Panel I: All rounds			
Effect of judgment (1 SD increase)	-0.015*		-0.024**
	(0.008)		(0.011)
Effect of cognitive ability (1 SD increase)		0.019*	0.026*
		(0.011)	(0.014)
Mean of dependent variable	0.05	0.05	0.05
Subject-round observations	999	999	999
Panel II: First round			
Effect of judgment (1 SD increase)	-0.002		-0.005
	(0.017)		(0.014)
Effect of cognitive ability (1 SD increase)		0.009	0.011
		(0.022)	(0.021)
Mean of dependent variable	0.04	0.04	0.04
Subject-round observations	141	141	141
Panel III: Rounds 2–10			
Effect of judgment (1 SD increase)	-0.018**		-0.027**
	(0.009)		(0.011)
Effect of cognitive ability (1 SD increase)		0.021	0.029*
		(0.012)	(0.015)
Mean of dependent variable	0.05	0.05	0.05
Subject-round observations	858	858	858

Notes: See the notes to Table 1. Panel I here replicates Panel I in Table 1. In Panels II and III here, we split the data by first round vs. rounds 2-10, instead of by first half vs. second half in Table 1.

Table A.3: Effects of cognitive skills on proportion of choices of zero:
First round vs. later rounds

	Chose one or two in round		
	(1)	(2)	(3)
Panel I: All rounds			
Effect of judgment (1 SD increase)	0.008		0.003
	(0.010)		(0.011)
Effect of cognitive ability (1 SD increase)		0.014*	0.013
		(0.008)	(0.009)
Mean of dependent variable	0.08	0.08	0.08
Subject-round observations	999	999	999
Panel II: First round			
Effect of judgment (1 SD increase)	0.005		0.005
	(0.005)		(0.005)
Effect of cognitive ability (1 SD increase)		0.001	-0.001
		(0.001)	(0.001)
Mean of dependent variable	0.01	0.01	0.01
Subject-round observations	141	141	141
Panel III: Rounds 2–10			
Effect of judgment (1 SD increase)	0.008		0.003
	(0.012)		(0.013)
Effect of cognitive ability (1 SD increase)		0.017*	0.016
		(0.010)	(0.011)
Mean of dependent variable	0.09	0.09	0.09
Subject-round observations	858	858	858

Notes: See the notes to Table A.1 in Web Appendix II. Panel I here replicates Panel I in Table A.1. In Panels II and III here, we split the data by first round vs. rounds 2-10, instead of by first half vs. second half in Table A.1.

Table A.4: Effects of cognitive skills on proportion of choices of one or two: First round vs. later rounds