How cognitive skills affect strategic behavior: Cognitive ability, fluid intelligence and judgment *

David Gill †

Zachary Knepper ‡

Victoria Prowse §

Junya Zhou ¶

This version: November 3, 2023

Abstract

We explore the influence of cognitive ability and judgment on strategic behavior in the beauty contest game. Using the level-k model of bounded rationality, cognitive ability and judgment both predict higher level strategic thinking. However, individuals with better judgment choose the Nash equilibrium action 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 the equilibrium action.

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

JEL classification: C92; C72; D91.

*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.

†Department of Economics, Purdue University; dgill.econ@gmail.com.

‡Department of Economics, Purdue University; zknepper@purdue.edu.

§Department of Economics, Purdue University; vprowse@purdue.edu.

¶Department of Economics, Purdue University; zhou830@purdue.edu.

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 (Oostrom et al., 2019), decision-making ability (Deming, 2021), problem solving (Autor, 2015), theory of mind (Fe et al., 2022) and creativity (Gill and Prowse, 2023a).

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.

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, 2023b).

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).

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:

- 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.[1] To measure judgment, subjects completed a test of situational judgment that measures judgment in work-related situations.[2] Practical intelligence and interpersonal skills help to underpin situational judgment ability (Stemler and Sternberg, 2006; Christian et al.,

[1]This test is similar to Raven's Progressive Matrices (Raven et al., 2000). Matrix reasoning tests have been used to study the effect of cognitive ability on strategic behavior in other settings by, e.g., Proto et al. (2019, 2022), Gill and Rosokha (2020) and Fe et al. (2022).

[2]Specifically, we used items that measure judgment in the domain of "analysis and problem solving" because that domain is most comparable to cognitive ability as measured by the matrix reasoning test. Oostrom et al. (2015) review situational judgment testing.

2010), and situational judgment scores predict job performance, after controlling for cognitive ability and job knowledge (Clevenger et al., 2001).

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 the Nash equilibrium action of zero, while subjects who are higher in judgment are *less likely* to choose the equilibrium action. 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 the equilibrium action 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 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 the Nash equilibrium action of 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 the equilibrium action 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 results, and Section 4 concludes. The Web Appendix provides further details.

2 Experimental design

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.

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). 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.

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 interpersonal work-related situations (see Web Appendix I.4 for a sample item). 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 (see Web Appendix I.5 for a sample item). Finally, subjects completed a short demographic questionnaire. Web Appendix I.3 provides further details.

As we explain in Web Appendix I.2, Online Appendix II.6 of Gill and Prowse (2023b) 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.

3 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 the Nash equilibrium choice of 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.[3]

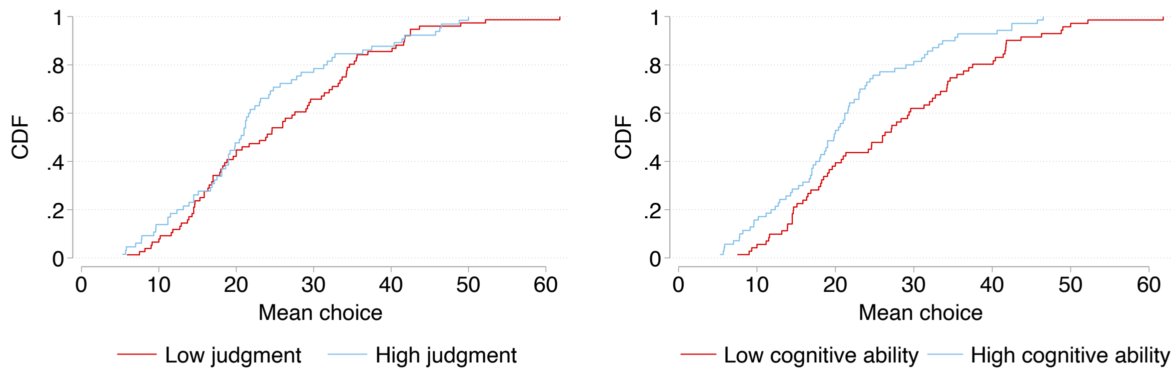
3.3 Reduced-form results

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 1 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 the Nash equilibrium choice of zero). The effects are strongest around the 75th percentile: at that percentile

[3]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 (2023a) find a correlation of 0.54 between cognitive ability and creativity.

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 1: CDFs of subject-level mean choice across rounds

From Figure 1, 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 the Nash equilibrium action of zero, while subjects who are higher in judgment are *less likely* to choose the equilibrium action. Since choices fall over rounds toward the Nash equilibrium choice of 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-equilibrium choices.

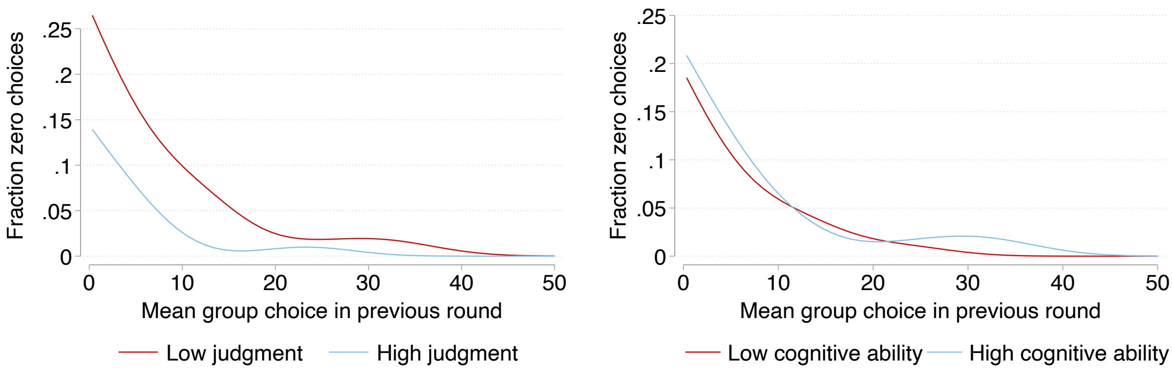
	Chose zero in round		
	(1)	(2)	(3)
Panel I: All rounds			
Effect of judgment (1 SD increase)	-0.015*		-0.024**
	(0.008)		(0.011)
Effect of cognitive ability (1 SD increase)		0.019*	0.026*
		(0.011)	(0.014)
Mean of dependent variable	0.05	0.05	0.05
Subject-round observations	999	999	999
Panel II: First 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

The result in Table 1 that subjects who are higher in judgment are less likely to choose the Nash equilibrium action of 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 the equilibrium action is driven by facets of judgment that are distinct from fluid intelligence.

We uncover a novel dynamic mechanism that helps to shed light on why higher judgment subjects are less likely to choose the Nash equilibrium action of zero. In particular, Figure 2 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.



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 1 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 the Nash equilibrium action) or strictly greater than 50.

Figure 2: Effect of mean group choice in previous round on proportion of choices of zero

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 (Stahl and Wilson, 1994; Nagel, 1995). 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 (2023b) 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 a subject's probability of being each level-k type, $Pr(k|\mathbf{z})$, depends on a vector $\mathbf{z}$ that includes measures of the subject's personal traits (including cognitive ability). A subject of type k then follows the noisy level-k choice rule in every round $r \geq 2$ (round 1 choices seed the model; equations (1), (2) and (3) in Section IV.A describe the level-k choice rules). Gill and Prowse (2016)'s 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.[4]

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-deviation increase in judgment increases the average level-k by around 0.15, which is a

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

9% increase relative to the average (specification (1); two-sided $p = 0.090$).[5] A one-standard-deviation increase in cognitive ability increases the average level-k by around 0.29 (specification (2); two-sided $p < 0.001$).[6] 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.

Taken together, these results indicate that fluid intelligence is a primary driver of strategic level-k thinking. This finding stands in contrast to our results from Section 3.3, which provided evidence that the lower inclination of high judgment subjects to choose the Nash equilibrium action of zero is driven by facets of judgment that are distinct from fluid intelligence.

[5] 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.

[6] 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 same structural level-k model as Gill and Prowse (2016), using the same maximum likelihood estimation routine, and using all 999 subject-round observations. Due to limitations imposed by the smaller number of observations here, starting from the model in Gill and Prowse (2016, Section V.C), we reduced the number of parameters by removing the level-4 type and the vector $\boldsymbol{\beta}$. Across specifications, we varied the contents of $\mathbf{z}$: in specification (1) we included the subject's score on the judgment test, in specification (2) we included the score on the cognitive ability test, and in specification (3) we included scores on both tests. Marginal effects were calculated by evaluating the type probabilities at the observed $\mathbf{z}$ for each subject and at the $\mathbf{z}$ with the relevant test score increased by one standard deviation (1 SD), taking the difference at the subject level, and then averaging across subjects. 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 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 the equilibrium action. 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 the Nash equilibrium action of 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

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

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

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

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

Christian, M.S., **Edwards, B.D.**, and **Bradley, J.C.** (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology*, 63(1): 83–117

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.** (2023a). The creativity premium: Exploring the link between childhood creativity and life outcomes. *Mimeo, Purdue University*

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

Gill, D. and **Rosokha, Y.** (2020). Beliefs, learning, and personality in the indefinitely repeated prisoner's dilemma. *CAGE Working Paper 489*

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

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

Oostrom, J.K., **de Vries, R.E.**, and **De Wit, M.** (2019). Development and validation of a HEXACO situational judgment test. *Human Performance*, 32(1): 1–29

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

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

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

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. At the end of each round, subjects received feedback about the group's choices in that round, 70% of the mean of the choices, and the earnings of the group members in that 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 (2023b) 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 (the other domains are: "working with others"; "resilience"; "delivering quality results"; "prioritising and organising"). 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 use 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.

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, 2020, also use the test in a setting with strategic behavior). Web Appendix I.5 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.

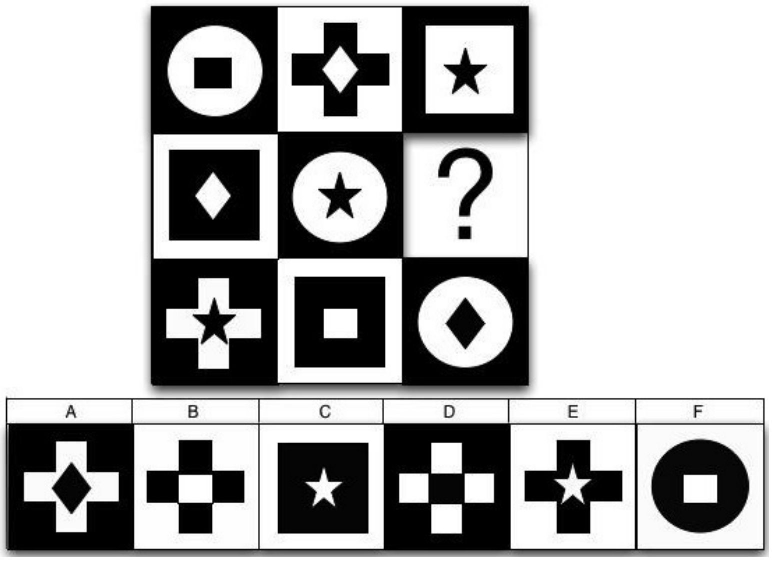
2. You go home and start your own research in order to remove the bugs issue and maintain the system. Your organisation has invested too much time and effort already for this change to cancel the replacement now.

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.

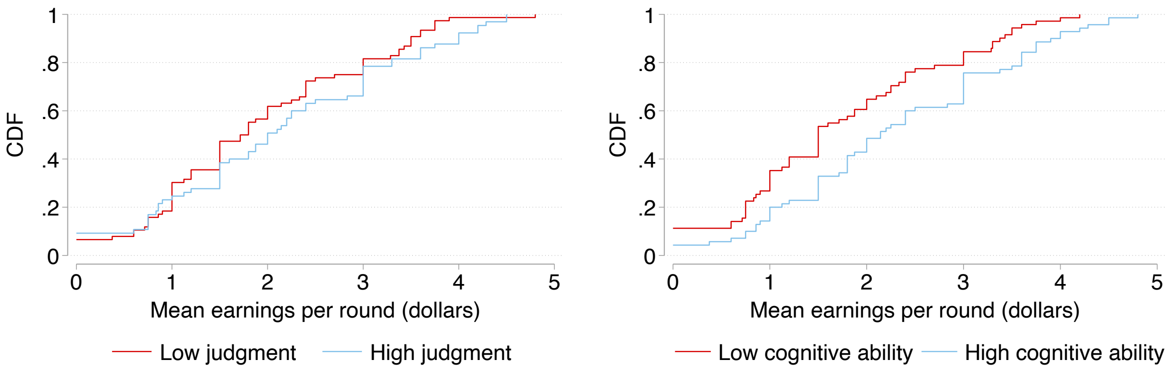
4. You speak with the staff members and use your persuasion skills to convince them that every choice made by the organisation is for the best, even if sometimes it does not end up being the most efficient one.

Web Appendix I.5 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 1 describe how we classify subjects by cognitive skill level.

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

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