Cognitive Ability, Character Skills, and Learning to Play Equilibrium: A Level-*k* Analysis

David Gill

Purdue University

Victoria Prowse

Purdue University

We investigate how cognitive ability and character skills influence the evolution of play toward Nash equilibrium in repeated strategic interactions. We find that more cognitively able subjects choose numbers closer to equilibrium, earn more, and converge more frequently to equilibrium play. We estimate a structural model of learning based on level *k* reasoning and find a positive relationship between cognitive ability and levels. Furthermore, the average level of more cognitively able subjects responds positively to the cognitive ability of their opponents. More agreeable and emotionally stable subjects also learn faster, although the effect of cognitive ability is stronger than that of personality.

I.   Introduction

In this paper, we aim to discover how cognitive ability and character skills influence behavior, success, and the evolution of play toward Nash equi-

librium in repeated strategic interactions.[1] Many economic interactions are strategic and repeated. Examples include sequences of market entry, exit and expansion decisions, repeated procuring or selling items via auctions, determining optimal pricing strategies in markets with dynamic price competition, engaging in repeated rounds of hiring or job searching, and competing within firms for promotions and bonuses. Given that cognitive ability and character skills vary greatly at the population level, to understand fully the dynamics of how people behave and learn in repeated strategic interactions, we need to establish the roles of cognition and character. Despite well-documented heterogeneity in cognitive ability and character skills, to the best of our knowledge we are the first to study how cognition and character affect how people learn to play equilibrium.

Our experiment was designed to answer the following questions: (i) Does cognitive ability influence the way in which groups of agents learn to play equilibrium over time when they engage in repeated strategic interaction? (ii) How do subjects' learning processes depend on both their own cognitive ability and that of their opponents? (iii) Do character skills influence how people learn to play equilibrium? (iv) How do the effects of character skills compare to those of cognitive ability?

We first measured the cognitive ability of our 780 subjects using a 60-question nonverbal Raven test. We classified each subject as either of "high cognitive ability" or of "low cognitive ability" according to whether her test score was in the top or bottom half of the distribution of scores in her session. We then matched our subjects into groups of three. In "own-matched" groups, all three members were of the same cognitive ability type. In "cross-matched" groups, the three members were of mixed ability (either two high-ability and one low-ability subject or vice versa). Subjects found out their own cognitive ability type as well as the ability type of the other two group members.[2] This design feature allows us to conduct an analysis of how behavior responds to the cognitive ability of opponents.

Each group of three then played the p beauty contest 10 times without rematching and with feedback.[3] In our p beauty contest, the three group

---

[1] We follow Heckman and Kautz (2014) in using the term "character skills" to describe personal attributes that are sometimes called "personality traits" or "noncognitive" skills. Heckman and Kautz note that character skills can be shaped and changed over the life cycle, while the term "traits" conveys "a sense of immutability or permanence."

[2] As far as possible the instructions used neutral language and so did not refer to "high ability," "low ability," "winning," and so on. Controlling for a subject's own Raven test score, we find no evidence that the allocation to cognitive ability type per se has any influence on choices or earnings (see Sec. A of the Appendix for details of the analysis).

[3] Real-world parallels to p beauty contests include timing games in financial markets: "Think of the time at which investors choose to sell a rising stock as picking a number. When many investors choose to sell, the stock crashes; the time of the crash is around the average number (selling time) chosen. Then professional investment is a p-beauty contest

members simultaneously chose an integer between 0 and 100 inclusive, and the subject whose choice was closest to 70 percent of the average of the three numbers won $6.00. In the unique Nash equilibrium, all subjects choose 0; however, the game is well suited to study learning since best responses to nonequilibrium choices are often above 0; but with repetition, behavior tends to move toward equilibrium. Our focus on repeated interaction makes it natural to keep group membership constant over the 10 rounds of the experiment; indeed, only with constant groups can we conduct our analysis of how often different groups converge to equilibrium play. We chose a group size of three to maximize the number of independent observations while ensuring that the game has an interesting strategic component (with a group size of just two, choosing 0 is weakly dominant).

We find striking differences by cognitive ability: more cognitively able subjects choose numbers closer to equilibrium, converge more frequently to equilibrium play, and earn substantially more on average even as behavior approaches the equilibrium prediction. In the final two rounds, all three group members choose the equilibrium action 29 percent of the time in own-matched high-ability groups but only 15 percent of the time in cross-matched groups and 13 percent of the time in own-matched low-ability groups; similar results hold for measures of close-to-equilibrium play.

To help gain insight into the microprocesses that drive these differences, and thus to understand better how subjects with different cognitive abilities learn differently, we estimate a structural model of learning based on level k reasoning that fits well the differences in behavior according to cognitive ability that we observe in the data.[4] Following Nagel

---

(with $p < 1$) in which investors want to sell a few days ahead of the crash—picking a number equal to p times the average number—but not too far ahead” (Ho, Camerer, and Weigelt 1998, 948).

[4] Methodologically, our structural analysis builds on existing level k mixture-of-types models estimated using maximum likelihood, including Stahl and Wilson (1995), Ho et al. (1998), Costa-Gomes, Crawford, and Broseta (2001), Costa-Gomes and Crawford (2006), Crawford and Iriberri (2007a, 2007b), and Costa-Gomes and Weizsäcker (2008). Structural level k mixture-of-types models have been applied successfully to study behavior in, e.g., guessing games (Costa-Gomes and Crawford 2006), coordination games (Costa-Gomes, Crawford, and Iriberri 2009), and auctions (Crawford and Iriberri 2007b) and to analyze the role of beliefs in normal-form games (Costa-Gomes and Weizsäcker 2008). The closely related cognitive hierarchy model (Camerer, Ho, and Chong 2004) has been used to study behavior in, e.g., zero-sum betting games (Brocas et al. 2014), telecom markets (Goldfarb and Xiao 2011), and the lowest unique positive integer game used by the Swedish national lottery (Östling et al. 2011). See Crawford, Costa-Gomes, and Iriberri (2013) for a comprehensive survey of applications of level k and cognitive hierarchy models. Our level k model of learning provides a relatively simple framework that succinctly organizes differences in behavior by cognitive ability. As is the case in all modeling exercises, we cannot rule out the existence of other models with superior or equal ability to fit our sample. Further, we cannot rule out the existence of alternative models that are observationally equivalent to our level k model of learning.

(1995), Stahl (1996), and Duffy and Nagel (1997), we assume that level 0 types "follow the crowd" in the sense that they copy the average group behavior from the previous round, level 1 types best respond to the choices of level 0 types, level 2 types best respond to the choices of level 1 types, and so on. We also incorporate a form of rule learning (Stahl 1996) by including types who switch up one level during the course of the experiment. The model allows subjects' levels to vary in their own cognitive ability and that of their opponents: we find a systematic positive relationship between cognitive ability and levels; furthermore, the average level of more cognitively able subjects responds positively to the cognitive ability of their opponents, while the average level of less cognitively able subjects does not respond.

The estimates of the structural model's parameters also allow us to simulate the earnings of each level k type, according to subjects' own cognitive ability and that of their opponents. This analysis shows that subjects are constrained in their levels below those that are optimal but that high–cognitive ability subjects leave less money on the table compared to the payoff-maximizing type (the payoff-maximizing type is allowed to vary in subjects' own cognitive ability and that of their opponents). On average, own-matched high-ability subjects leave $4.59 on the table, cross-matched high-ability subjects leave $4.61 on the table, own-matched low-ability subjects leave $5.75 on the table, and cross-matched low-ability subjects leave $7.75 on the table.

Before the test of cognitive ability, we administered questionnaires to 270 of the 780 subjects to measure seven character skills: openness, conscientiousness, extraversion, agreeableness, emotional stability (the Big Five), grit, and consideration of future of consequences (CFC).[5] The high degree of correlation between our seven measures justifies the construction of a smaller number of uncorrelated factors. A principal component factor analysis gives three factors, which explain two-thirds of the variance: factor 1 mainly captures conscientiousness, grit, and CFC; factor 2 mainly captures agreeableness and emotional stability; while factor 3 mainly captures openness, extraversion, and CFC.

We find that both cognitive ability and character skills influence behavior, success, and learning in our repeated strategic interaction. Subjects with a higher Raven test score (who are more cognitively able) and a higher factor 2 score (who are more agreeable and emotionally stable) choose numbers closer to equilibrium, earn more, and, within the context of our level k model of learning, adopt higher level k choice rules. We also find that both cognitive ability and character skills influence learning at the

---

[5] We measured the Big Five using the Big Five Inventory (John, Donahue, and Kentle 1991; John, Naumann, and Soto 2008). The Grit Scale measures "perseverance and passion for long-term goals" (Duckworth et al. 2007). CFC is a survey-based measure of time preference (Strathman et al. 1994).

group level: groups with higher cognitive ability on average and higher factor 2 on average end up closer to the Nash equilibrium in the last few rounds. Recent evidence from the psychology literature links agreeableness to a better theory of mind (Nettle and Liddle 2008; DeYoung et al. 2010),[6] thus providing a channel through which agreeableness might help performance in settings like ours that require strategic reasoning, and shows that emotional stability aids performance in more complex tasks (Eysenck 1985; Szymura and Wodniecka 2003).

The impact of cognitive ability on behavior is about 75 percent bigger than that of character skills, but the estimated influence of cognitive ability does not depend on whether we control for character skills and vice versa. Interestingly, we also find that marginal changes in cognitive ability mainly affect the behavior of subjects who are already cognitively able, while changes in character skills influence the behavior of subjects who are high or low in cognitive ability. Our results mirror those of recent studies that show that both cognitive ability and character skills help to predict outcomes such as workplace performance, educational attainment, marital status, and risky behaviors (Heckman, Stixrud, and Urzua 2006; Mueller and Plug 2006; Duckworth et al. 2007; Borghans et al. 2008; Cunha, Heckman, and Schennach 2010; Lindqvist and Vestman 2011; Burks et al. 2014; Heckman, Humphries, and Veramendi 2015). Consonant with our finding that marginal increases in cognitive ability matter more for subjects who are already cognitively able, Lindqvist and Vestman (2011) find that noncognitive skills predict wages more strongly than do cognitive skills at the lower end of the earnings distribution, while cognitive skills are the stronger predictor at the higher end of the earnings distribution. Heckman et al. (2006) and Heckman et al. (2015) show that the relative importance of cognitive and noncognitive skills varies according to the outcome: for example, Heckman et al. (2006) find that cognitive ability has the same effect on wages as noncognitive ability, while noncognitive ability is the dominant factor in explaining participation in crime (see the surveys by Almlund et al. [2011] and Heckman and Kautz [2012] for further examples). An interesting extension of our study would investigate the relative importance of cognition and character in explaining learning in different strategic games or in other tasks. It would be particularly interesting to study whether the factor intensities attached to cognitive and noncognitive skills vary between games in which subjects learn a less-than-obvious dominant strategy, such as race games (Dufwenberg, Sundaram, and Butler

---

[6] According to Coricelli and Nagel (2009), theory of mind is "the ability to think about others' thoughts and mental states to predict their intentions and actions" (9163) and playing the beauty contest against humans rather than a computer "activated areas commonly associated with theory of mind or mentalizing-thinking about other people's minds" (9166).

2010; Gneezy, Rustichini, and Vostroknutov 2010; Cardella 2012) and games like the beauty contest in which players do not have a dominant strategy.

The contribution of our paper is to investigate how cognitive ability and character skills influence learning to play equilibrium in repeated strategic interactions. In doing so, we extend three established literatures. The first literature studies learning in strategic games, both theoretically and empirically, without considering the influence of cognitive ability or character skills (see Camerer [2003] for a survey). The second literature studies the relationship between cognitive ability and behavior in strategic games without considering the dynamics of how people learn over time to play equilibrium.[7] The third literature studies the relationship between character skills and behavior in strategic games, again without considering how people learn to play equilibrium.[8]

A further innovation of our paper is to study how learning behavior varies in the cognitive ability of opponents and to study how subjects' own cognitive ability influences how they respond to their opponents' cognitive ability. This analysis extends a burgeoning literature that studies how strategic play is affected by information about opponent characteristics other than cognitive ability. In particular, a small literature looks at the effects of information about how much experience opponents have of the type of strategic reasoning required in the game being studied. Notably, Agra-

---

[7] In one-shot beauty contest games, Burnham et al. (2009) and Brañas-Garza, García-Muño, and Hernán (2012) find that subjects with higher cognitive ability choose lower numbers, while Agranov, Caplin, and Tergiman (2011) find that, when subjects are given time to think about their choices, higher–cognitive ability subjects' choices fall more with thinking time. In related dominance-solvable and guessing games, working memory (Rydval, Ortmann, and Ostatnicky 2009) and depths of reasoning in the red hat puzzle (Bayer and Renou 2012) correlate with behavior. Cognitive ability also influences behavior in public-good games (Millet and Dewitte 2007) and in the prisoner's dilemma (Burks et al. 2009). Schnusenberg and Gallo (2011) run a three-round beauty contest with one winner per round and no monetary incentives (the prize was a small in-class grade improvement) and find that scores in Frederick's (2005) three-question cognitive reflection test affect choices in the first round only.

[8] In public-good games, self-esteem (Kurzban and Houser 2001) and agreeableness (Volk, Thöni, and Ruigrok 2011) generate higher contributions, and more agreeable subjects choose to lead less frequently (Arbak and Villeval 2013). More self-controlled, more tough-minded, and less extraverted subjects offer more in the ultimatum game (Brandstätter and Königstein 2001). More agreeable subjects send more in a sequential prisoner's dilemma (Rustichini et al. 2012). Higher need for cognition and premeditation give rise to fewer dominated choices in dominance-solvable games (Rydval et al. 2009). More neurotic subjects earn more in the stag hunt game (Al-Ubaydli, Jones, and Weel 2013). In trust games, less anxious (Fahr and Irlenbusch 2008), less alienated (Johnson, Rustichini, and MacDonald 2009), less conscientious, less neurotic, and more agreeable (Müller and Schwieren 2012) subjects are more trusting. Less anxious and more aggressive subjects are more likely to make a risky entry decision (Neeman, Kugler, and Vulkan 2013). This literature does not focus on how character skills influence learning to play equilibrium, although Boone et al. (2002) find that sufficient repetition allows subjects with an external locus of control to become more cooperative in the prisoner's dilemma.

nov et al. (2012) find that, in a one-shot beauty contest, inexperienced undergraduates shift to higher level k types on average when they play against graduates with some experience of the game, while Le Coq and Sturluson (2012) find that play in a Cournot game shifts with the experience of opponents in the game, and Palacios-Huerta and Volij (2009) find an effect of opponent experience of backward-induction reasoning on behavior in the centipede game.

We study how cognitive ability measured in a nonstrategic setting affects how people perform and learn in a strategic environment. In contrast, Palacios-Huerta and Volij (2009) focus on the role of experience of a particular type of strategic reasoning (backward induction) on how subjects play in a repeated strategic game that uses this type of reasoning (the centipede game). To vary the amount of experience, Palacios-Huerta and Volij compared students and chess players in a laboratory experiment. As the authors state, "Backward induction reasoning is second nature to expert chess players. They devote a large part of their life to finding optimal strategies for innumerable chess positions using this reasoning" (1624). The chess players were recruited from local chess clubs, and none held an official chess title; thus there is no evidence that they were of exceptional intelligence.[9] Palacios-Huerta and Volij did not measure cognitive ability, and they do not claim that the chess players were more intelligent than the students. Even if we accept that the chess players were more intelligent (on one or more dimensions) as well as more experienced, Palacios-Huerta and Volij's data cannot be used to distinguish the effects of cognitive ability on behavior from that of experience: a study that aims to use chess playing as a proxy for intelligence should use a game that does not require backward-induction reasoning.

The paper proceeds as follows. Section II describes the experimental design. Section III reports descriptive statistics and reduced-form regression results that describe how behavior and learning vary with cognitive ability. Section IV presents the structural analysis of the influence of cognitive ability. Section V compares the effects of cognitive ability to those of character skills. Section VI discusses some broader implications of our findings and presents conclusions.

---

[9] In a separate field experiment reported in Palacios-Huerta and Volij (2009), chess players played the centipede game a single time. These chess players were recruited from international chess competitions and included a number of grand masters, international masters, and federation masters. Palacios-Huerta and Volij find that 69 percent of first movers stopped at the first node, with the proportion rising to 100 percent for grand masters. In a close replication (with the same game and information about the opponent, and with similar subjects and stakes), Levitt, List, and Sadoff (2011) find different results: only 4 percent of first movers stopped at the first node, with the proportion falling to 0 percent for grand masters.

II.   Experimental Design

We ran 37 experimental sessions at the University of Arizona's Experimental Science Laboratory (ESL), all conducted on weekdays between November 2010 and March 2011 and in January 2014 and lasting approximately 75 minutes. Eighteen or 24 student subjects participated in each session, with 780 participants in total.[10] The participants were drawn from the ESL subject pool, which is managed using a bespoke online recruitment system. We excluded any graduate students in economics. Seating positions were randomized. The experimental instructions (see the supplementary web appendix) were provided to each subject on his or her computer screen and were read aloud to the subjects. Questions were answered privately. Each subject was paid a show-up fee of $5.00 and earned an average of a further $20.00 during the experiment (all payments were in US dollars). Subjects were paid privately in cash. The experiment was programmed in z-Tree (Fischbacher 2007).

A.   The Test of Cognitive Ability

Each session consisted of a test of cognitive ability followed by repeated play of the p beauty contest (Nagel 1995). (The sessions that we ran in January 2014 also included questionnaires to measure character skills that were administered before the test of cognitive ability; Sec. V provides the details of the measures and analysis of the impact of character skills.) In more detail, each session included a 30-minute computerized test of cognitive ability using Raven's Progressive Matrices. The Raven test, which consists of nonverbal multiple-choice questions, is recognized as a leading measure of analytic intelligence (Carpenter, Just, and Shell 1990; Gray and Thompson 2004, 472, box 1). [11] In economics, Raven test scores have been found to correlate positively with fewer Bayesian updating errors (Charness, Rustichini, and van de Ven 2011) and with more accurate beliefs (Burks et al. 2009). Each question asks the subjects to identify the missing element that completes a visual pattern. We used the Stan-

---

[10] All sessions were run during the fall or spring semesters. Before running these sessions, we also ran one pilot session without any monetary incentives and two sessions with a different form of the p beauty contest (where the target was 90 percent of the mean of the choices) whose results are not reported here.

[11] Carpenter et al. (1990) define analytic intelligence (also sometimes called fluid intelligence, as opposed to crystallized intelligence) as "the ability to reason and solve problems involving new information, without relying extensively on an explicit base of declarative knowledge derived from either schooling or previous experience" (404), and they show that Raven test scores discriminate according to the ability to use abstract reasoning and correlate highly with scores on other complex cognitive tasks. Raven, Raven, and Court (2000, SPM25–SPM37) survey the extensive literature that studies (i) correlations between Raven test scores and scholastic aptitude and achievement tests and other measures of cognitive ability and (ii) the internal consistency and test-retest reliability of the Raven test.

dard Progressive Matrices Plus version of the Raven test (the level of difficulty lies between that of the Standard Progressive Matrices and the Advanced Progressive Matrices), which consists of 60 questions split into five parts of increasing difficulty, labeled A–E, with 12 questions in each. We gave the subjects 3 minutes for each of parts A and B (which are easier than parts C–E) and 8 minutes for each of parts C, D, and E. Within each part subjects could move back and forth between the 12 questions in that part, and time permitting, they were allowed to change their previous answers. The subjects were not made aware of what would follow the Raven test until the test was completed.

We did not provide any monetary incentives for completing the Raven test. This is conventional in the psychology and psychometric literatures and avoids the possibility that income effects might spill over from the test to behavior in the p beauty contest. We did, however, tell the subjects that we would inform them privately of their own score at the very end of the session. Figure 1A shows how the Raven test scores of our subjects were distributed. The mean test score was 40.7, with individual scores ranging from 12 to 58.

B. Subject Matching

Following the Raven test, each subject was determined to be either (i) of "high cognitive ability" if her test score was in the top half of the distribution of scores of the subjects in her session or (ii) of "low cognitive ability" if her test score was in the bottom half of the distribution in her session. Subjects scoring exactly the session median were allocated to ability types so as to ensure an equal number of subjects of each ability type for that session. Figure 1B illustrates the densities of the Raven test scores by cognitive ability type. On average, the Raven test score of high-ability subjects was 7.0 higher than that of low-ability subjects. The range of overlap arises because the median Raven test score varied over sessions, from 37.0 to 43.5. We found no time trend in the median test score over sessions (two-sided $p = .530$).

Subjects were then put into groups of three. In "own-matched" sessions, all three members of a group were of the same ability type. In own-matched sessions, 450 subjects participated, giving 75 groups of three high–cognitive ability subjects ("own-matched high-ability groups") and 75 groups of three low–cognitive ability subjects ("own-matched low-ability groups"). In "cross-matched" sessions, the three members of a group were of mixed ability (half the groups were made up of two high–cognitive ability subjects matched with one low–cognitive ability subject and the other half were made up of two low–cognitive ability subjects matched with one high–cognitive ability subject). In cross-matched sessions, 330 subjects participated, giving 110 cross-matched groups. Conditional on a subject's cog-
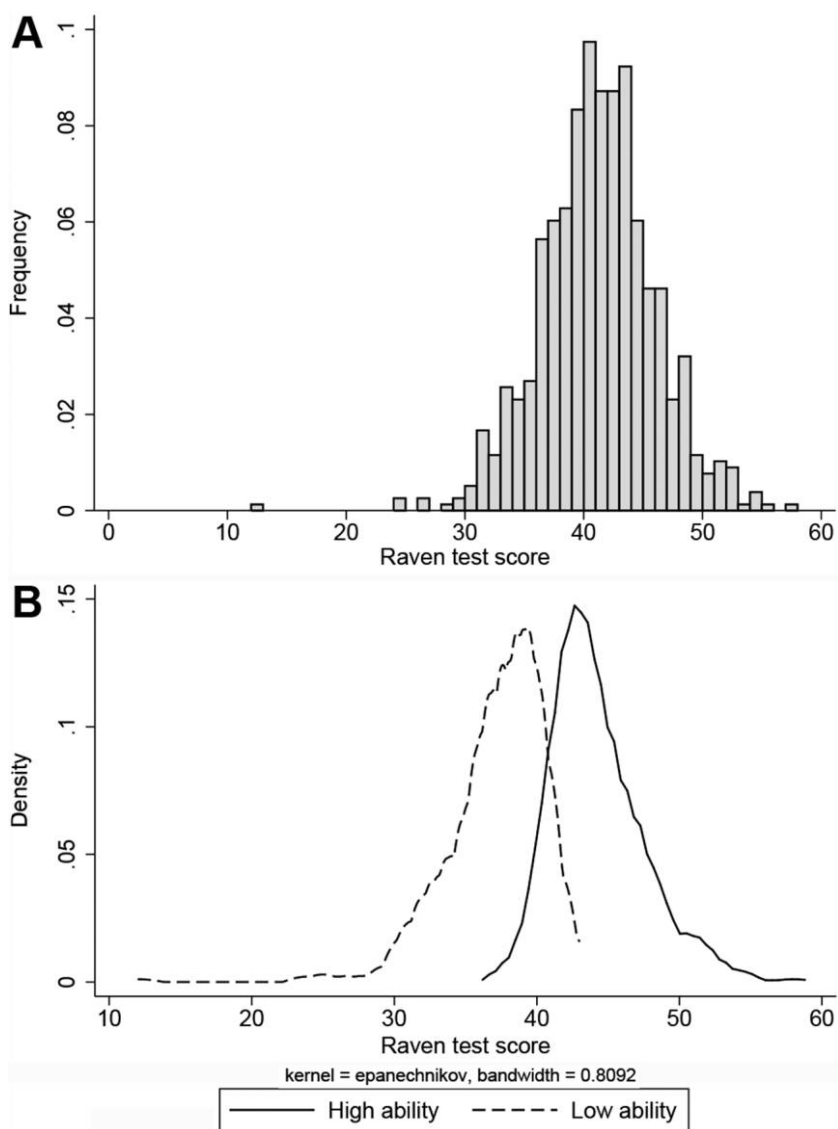
FIG. 1.—Histogram and densities of Raven test scores. *A*, Histogram of Raven test scores. *B*, Smoothed densities of Raven test scores.

nitive ability type, the allocation to groups was random. The subject matching implies that, on average, high-ability subjects face four-thirds more high-ability opponents in own-matched groups than in cross-matched groups, while low-ability subjects face four-thirds more high-ability opponents in cross-matched groups than in own-matched groups.

We informed each subject of her own cognitive ability type as well as the cognitive ability type of the two other members of her group. The instructions (see the supplementary web appendix) did not use the terms "high ability" or "low ability"; instead, we referred more neutrally to the top and bottom halves of the test scores of all participants in the room. Controlling for a subject's own test score, we find no evidence that the allocation to cognitive ability type per se influenced behavior or earnings in the experiment: Section A in the Appendix provides the details of this analysis.

C. p Beauty Contest Game

Each group of three then played 10 rounds of the p beauty contest (Nagel 1995) with $p = .7$ and without rematching (no rematching allows us to treat behavior across groups as independent). In particular, in every round each group member privately chose an integer between 0 and 100 inclusive (the subjects typed their chosen number into a box rather than selecting it from an on-screen grid). The group member whose chosen number was closest to 70 percent of the mean of the three numbers chosen by the group members (the "target") was paid $6.00 and the other group members received nothing. In the case of ties, the $6.00 was split equally among the winners. To keep the language as neutral as possible, the instructions (see the supplementary web appendix) did not use terms such as "prize," "winner," "loser," "ties," or "target." The unique Nash equilibrium is for all players to choose 0.[12]

Before the start of the first round, we described the number of rounds, the rules of the game, and the information feedback the subjects would receive at the end of each round. At the end of every round, each group member was informed of (i) the numbers chosen by the group members, (ii) the mean of the three chosen numbers, (iii) 70 percent of the mean (the target), (iv) which group member's (members') number(s) was (were) closest to the target, and (v) how much each group member

---

[12] The game is discrete, and hence this is not true for all values of p (López 2001). However, it is relatively straightforward to show for our $p = .7$. A proof starts from the observation that the highest chosen number can never win or tie unless all three players choose that number. Suppose that $x_i \geq x_j$ and $x_i > x_h$. Let t be the target. If $x_h \geq t$, clearly i cannot win or tie. If $x_h < t$, x_h is closer to the target than x_i if and only if $(x_i - t) - (t - x_h) > 0$. When $p = .7$, this difference equals $(1/15)(8x_i - 7x_j + 8x_h) > 0$. Now suppose that a Nash equilibrium exists in which the players do not always all choose 0. Let x^{max} be the highest number that is ever played in equilibrium; x^{max} cannot be a best response. From the observation, the payoff from $x_i = x^{max}$ is strictly positive if and only if $x_j = x_h = x^{max}$ (giving a tie). If that can happen with strictly positive probability, deviating to any lower number is profitable (from the observation giving a win when $x_j = x_h = x^{max}$). If not, deviating to the lowest number ever chosen by your opponents is profitable (from the observation giving a tie when at least one opponent chooses that number).

was paid for the round.[13] While deciding on their choice of number and also when receiving feedback, the subjects could see a reminder of the rules and of the cognitive ability type of each member of their group. All interactions were anonymous, but subjects were given labels (X, Y, or Z), which were held fixed for the 10 rounds; hence subjects could link the choices in their group to particular opponents whose cognitive ability type they knew.

The subjects had 90 seconds to make their choice in each round. The subjects were told that if they made their choice early, they would still have to wait for the full 90 seconds. If a subject failed to make a choice within 90 seconds, a flashing request prompted an immediate choice. At the end of each round, the subjects could see the feedback information described above for a period of 30 seconds before the next round began.

III.   Reduced-Form Results

In this section, we report descriptive statistics and reduced-form regression results in order to describe how average behavior, the evolution of play over rounds, and the group-by-group dynamics of convergence toward equilibrium vary with cognitive ability. In Section IV, we go on to build and estimate a structural model of learning that aims to provide insight into the behavioral mechanisms that underlie the patterns that we describe below.

A.   *Behavior and Earnings in the First Round*

We start by reporting briefly how behavior and earnings vary with cognitive ability in the first round. The mean choice of high–cognitive ability subjects is 42.9 while that of low–cognitive ability subjects is 43.6. The difference of 0.7 is not statistically significant (a regression of p beauty contest choices on cognitive ability type gives a two-sided $p = .626$).[14] To study differences in earnings by cognitive ability we look at subjects only in cross-matched groups: low–cognitive ability subjects in own-matched low-ability groups by construction must earn as much on average as high–

---

[13] In terms of information feedback, our design is close to that of Nagel (1995), where the whole distribution of choices was revealed at the end of each round (but with much larger groups and just four repetitions).

[14] Throughout Secs. III.A and III.B, all statistical tests are based on ordinary least squares (OLS) regressions and use heteroskedasticity-consistent standard errors. Apart from the case of first-round choices, we cluster at the group level to allow for within-group nonindependence. In Sec. III.A the regressions use the 780 subjects observed for one round, in Sec. III.B.1 they use the 780 subjects observed for 10 rounds, in Sec. III.B.2 they use the 450 own-matched subjects observed for either five or 10 rounds, and in Sec. III.B.3 they use the 330 cross-matched subjects observed for either five or 10 rounds.

cognitive ability subjects in own-matched high-ability groups (i.e., $2.00 on average in every round). In the first round, cross-matched high-ability subjects earn $0.11 more than cross-matched low-ability subjects (mean earnings in the first round are $2.00), but the difference is not statistically significant (two-sided $p = .761$).

B.   Behavior across All 10 Rounds

We now consider how behavior varies with cognitive ability across all rounds of the experiment, describing both average behavior across all 10 rounds and the evolution of behavior during the course of the experiment. We first study the behavior of all subjects and then focus on own-matched subjects and cross-matched subjects separately.[15] We study differences in earnings only for cross-matched subjects: as noted in Section III.A, low–cognitive ability subjects in own-matched low-ability groups by construction must earn as much on average as high–cognitive ability subjects in own-matched high-ability groups (i.e., $2.00 on average in every round).

1.   Behavior of All Subjects

Across all 10 rounds of the p beauty contest, the mean choice of high–cognitive ability subjects is 17.9 while that of low–cognitive ability subjects is 20.1. Regressing p beauty contest choices on cognitive ability type, we find that the difference of 2.2 is highly statistically significant (two-sided $p = .007$). Figure 2 shows the round-by-round evolution of mean choices for high– and low–cognitive ability subjects. We can see that average behavior moves toward the Nash equilibrium for both types. Shortly, we will show that the high-ability subjects earn significantly more than the low-ability subjects (Sec. III.B.3) and converge to a greater degree to Nash equilibrium (Sec. III.C.2).

2.   Behavior of Own-Matched Subjects

The difference in average behavior between high– and low–cognitive ability subjects is more pronounced for own-matched subjects than for the sample as a whole. Across all 10 rounds, the mean choice of own-matched high–cognitive ability subjects is 3.5 lower than that of own-matched low-ability subjects (two-sided $p = .005$). Figure 3 shows round-by-round mean choices for own-matched subjects only. In rounds 1–5, choices of own-matched high–cognitive ability subjects are on average 3.1 lower than choices of own-matched low-ability subjects (two-sided $p = .025$).

---

[15]   Sections C and D of the Appendix provide further analysis of the cross-matched subjects.
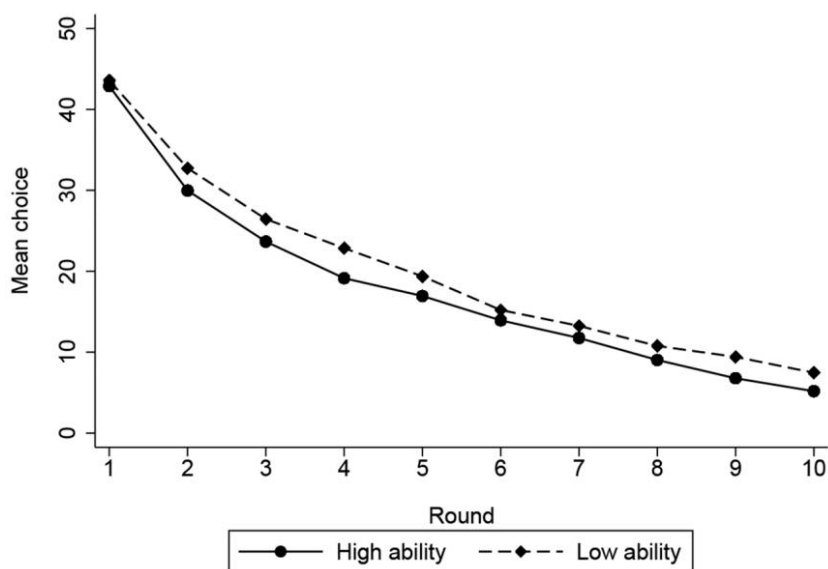
Fig. 2.—Round-by-round means of p beauty contest choices (all subjects)

In rounds 6–10, own-matched high–cognitive ability subjects' choices remain lower, by an average of 3.9 (two-sided $p = .006$). By the final round, the mean choice of own-matched high–cognitive ability subjects falls to 4.3, while that of own-matched low-ability subjects falls to 8.2.
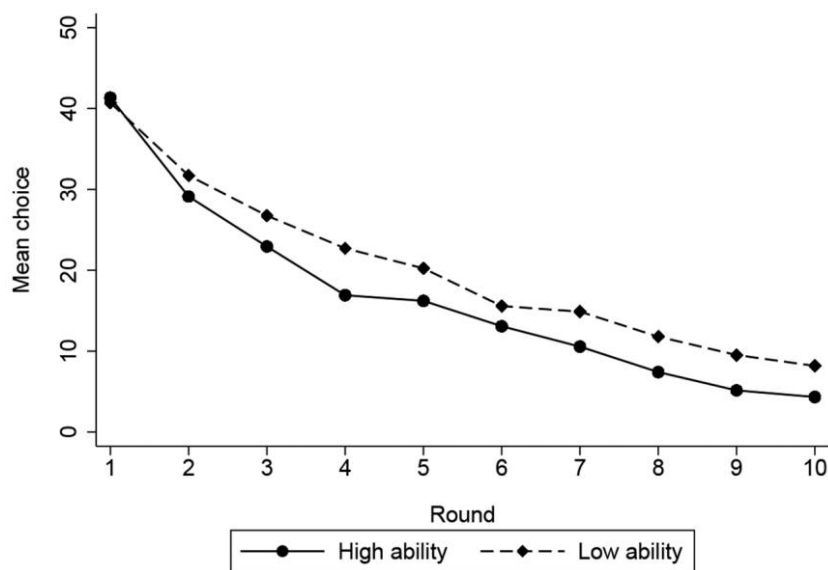
Fig. 3.—Round-by-round means of p beauty contest choices of own-matched subjects

3.   Behavior and Earnings of Cross-Matched Subjects

Across all 10 rounds, the mean choice of cross-matched high–cognitive ability subjects is only 0.4 lower than the mean choice of cross-matched low-ability subjects, and the difference is not statistically significant (two-sided $p = .653$). Figure 4A shows round-by-round mean choices for cross-
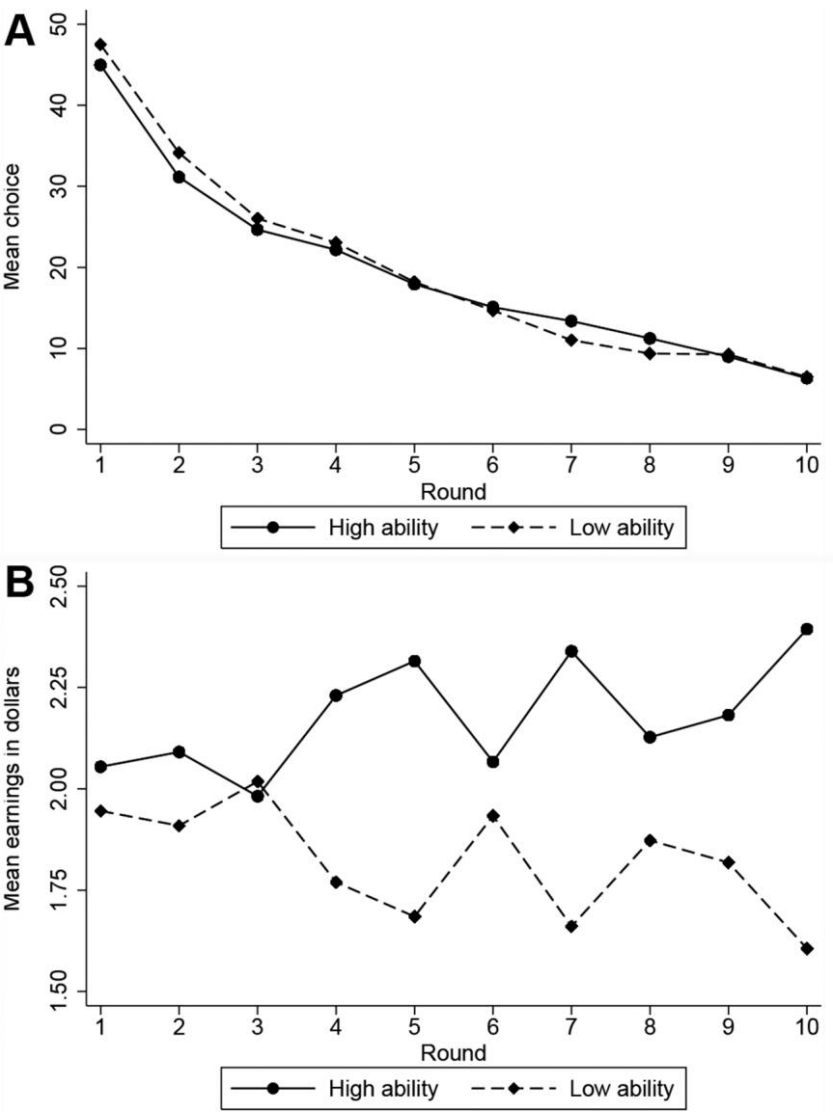
FIG. 4.—p beauty contest choices and earnings of cross-matched subjects. A, Round-by-round means of choices. B, Round-by-round means of earnings.

matched subjects only. Cross-matched high– and low–cognitive ability subjects behave similarly on average in both the first and second halves of the experiment: there is no statistically significant difference in behavior in either rounds 1–5 (two-sided $p = .181$) or rounds 6–10 (two-sided $p = .330$).

Despite the similarity in the average behavior of cross-matched high– and low–cognitive ability subjects, we find that cross-matched high–cognitive ability subjects are more successful in the sense that they earn significantly more money in the experiment. Over the 10 rounds, cross-matched high-ability subjects earn $3.56 more than cross-matched low-ability subjects, and the difference is highly statistically significant (two-sided $p = .007$).

Figure 4*B* shows how the earnings of cross-matched subjects evolve during the course of the experiment. We can see that (i) cross-matched high-ability subjects earn more than cross-matched low-ability subjects in both the first and second halves of the experiment and that (ii) the earnings difference becomes bigger in the second half. In rounds 1–5, cross-matched high–cognitive ability subjects earn $0.27 more per round than cross-matched low-ability subjects, although the difference is not quite statistically significant (two-sided $p = .143$). In rounds 6–10, cross-matched high–cognitive ability subjects earn $0.44 more per round than cross-matched low-ability subjects, and the difference is highly statistically significant (two-sided $p = .003$). By the final round, the difference in earnings rises to $0.79.

Thus earnings diverge in the second half of the experiment, even though average behavior remains similar. Somehow, cross-matched high–cognitive ability subjects "learn" better how to play the game than do cross-matched low–cognitive ability subjects. An important aim of the structural analysis described in Section IV is to provide insight into the mechanisms driving these results.

C.   *Group-by-Group Convergence toward Equilibrium*

Section III.B provides a good overview of how average behavior evolves over time. However, this overview masks a significant amount of group-by-group variation in exactly how play evolves toward equilibrium. In order to better understand how the dynamics of the learning process vary with cognitive ability, in this section we look in more detail at the process of convergence toward Nash equilibrium. We start by providing a visual description of the way in which some individual groups succeed or fail to converge toward equilibrium; we then present results about the proportion of groups that converge and how the degree of convergence varies with group composition.

1.   Visual Description of Group-by-Group Behavior

Figures 5 and 6 give an overview of the group-by-group variation in the dynamics of convergence. Each of the eight figures shows, for a specific group, how the choices of the three group members change round by round. The figures also show the winning choice or choices in each round.[16]

Figures 5*A*–5*C* show examples of three groups that slowly converge toward equilibrium. No group ever reaches equilibrium in figures 5*A*–5*C*. However, some groups do successfully converge all the way to equilibrium. Figure 5*D* shows an own-matched high-ability group in which convergence to equilibrium is almost immediate: by the fourth round all three group members choose 0, and all three then stick to the equilibrium choice for the remainder of the experiment. Figure 6*A* shows a cross-matched group in which behavior also converges to equilibrium, although not as fast. As we will see shortly in Section III.C.2, convergence is much more common when all three members of the group are of high cognitive ability.

Convergence toward equilibrium is not the only pattern that we observe. Figures 6*B* and 6*C* show groups in which behavior does not move discernibly toward equilibrium. In both cases, high choices by some of the subjects seem to disturb the learning process. However, figure 6*D* shows an own-matched high-ability group in which such a high choice seems hardly to affect the learning process at all.

2.   Statistics on Group-by-Group Convergence

Of course, the figures in Section III.C.1 illustrate the behavior of just a small subset of the 260 groups in our sample. We now study the degree of convergence more systematically, differentiating between own-matched high-ability groups, cross-matched groups, and own-matched low-ability groups. Table 1 shows the frequency of equilibrium and close-to-equilibrium play in the final two rounds of the experiment.

Column 1 shows the proportion of equilibrium play, that is, the proportion of group round observations in which all three group members choose 0. Own-matched high-ability groups play the equilibrium around 29 percent of the time in the final two rounds, which is statistically significantly more often than for cross-matched groups (15 percent) and own-matched low-ability groups (13 percent). Columns 2 and 3 show that the results extend when we consider close-to-equilibrium play, defined to be

---

[16] The supplementary web appendix contains figures illustrating the dynamics of convergence for each of the 260 groups in our sample.
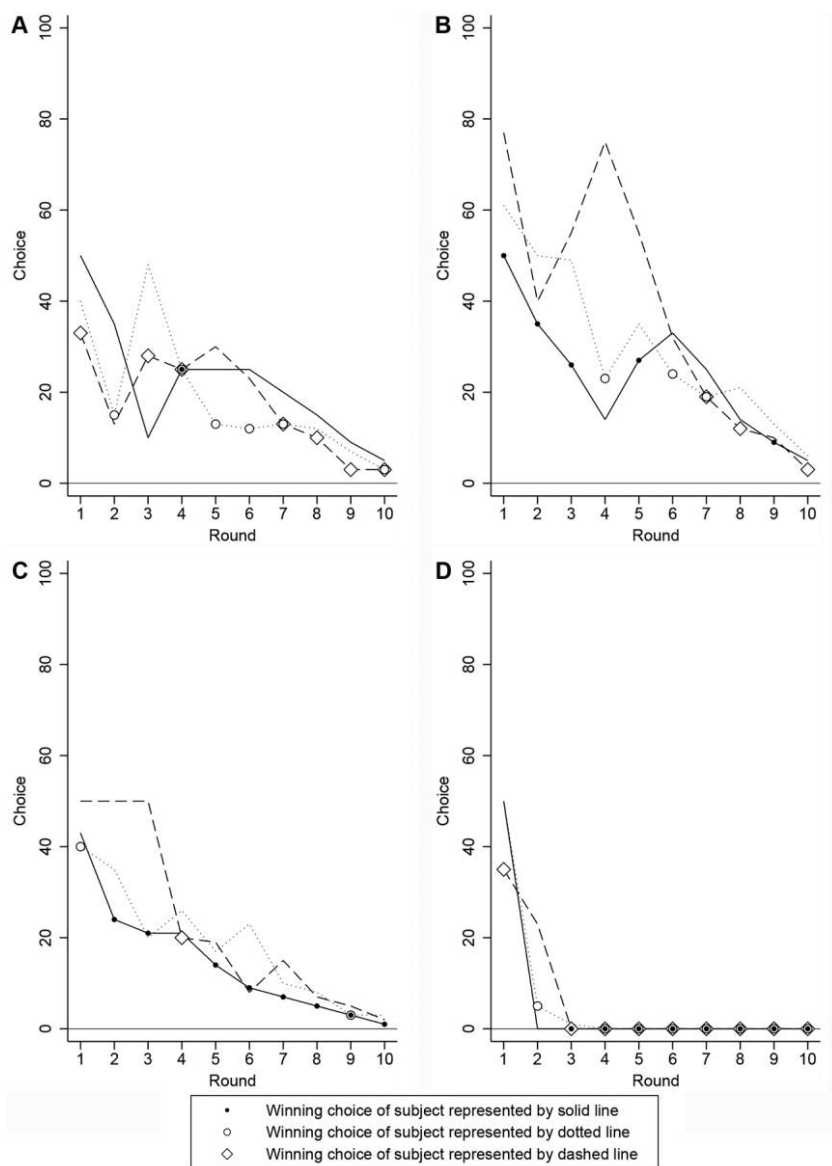
FIG. 5.—Examples of group-by-group behavior: groups 1–4 of eight. *A*, Own-matched high-ability group. *B*, Own-matched low-ability group. *C*, Cross-matched group with one high-ability subject. *D*, Own-matched high-ability group. Ties are represented by overlapping markers. In cross-matched groups, the subject represented by the solid line is of the minority cognitive ability type.

1636

Fɪɢ. 6.—Examples of group-by-group behavior: groups 5–8 of eight. *A*, Cross-matched group with one high-ability subject. *B*, Cross-matched group with two high-ability subjects. *C*, Own-matched low-ability group. *D*, Own-matched high-ability group. Ties are represented by overlapping markers. In cross-matched groups, the subject represented by the solid line is of the minority cognitive ability type.
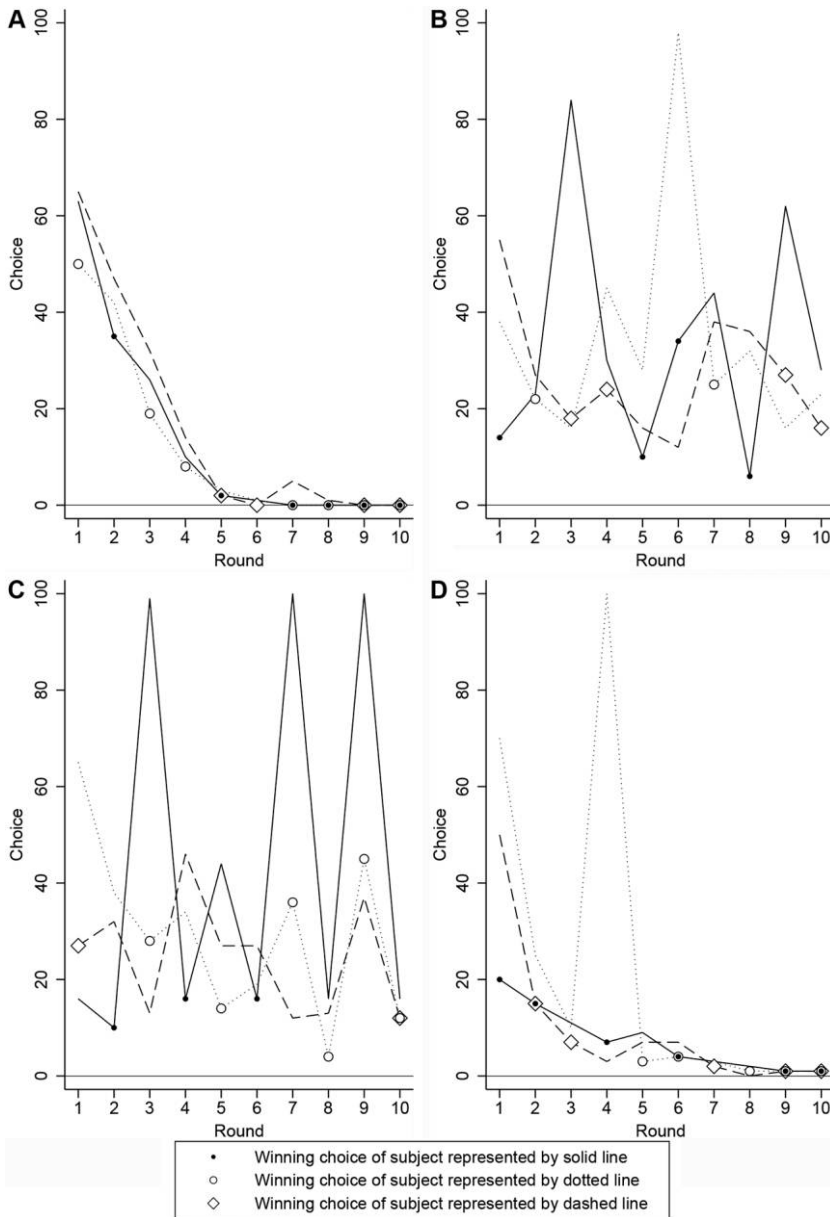
1637

TABLE 1
PROPORTIONS OF EQUILIBRIUM AND CLOSE-TO-EQUILIBRIUM PLAY IN ROUNDS 9 AND 10

	Equilibrium (1)	Group Mean ≤1 (2)	Group Mean ≤2 (3)
Observed proportions:			
Own-matched high-ability groups	.293	.453	.540
Cross-matched groups	.145	.282	.368
Own-matched low-ability groups	.127	.247	.313
Cross-group differences in proportions:			
Own-matched high-ability groups vs.			
own-matched low-ability groups	.167***	.207***	.227***
	(.062)	(.072)	(.072)
Own-matched high-ability groups vs.			
cross-matched groups	.148**	.172**	.172**
	(.059)	(.069)	(.070)
Cross-matched groups vs. own-matched			
low-ability groups	.019	.035	.055
	(.049)	(.061)	(.064)

NOTE.—The table is based on 520 group-round observations (the 75 own-matched high-ability groups, the 75 own-matched low-ability groups, and the 110 cross-matched groups, observed in rounds 9 and 10). Column 1 reports the proportion of group-round observations in rounds 9 and 10 in which all three group members choose 0. Column 2 (3) reports the proportion of group-round observations in rounds 9 and 10 in which the mean choice of the three group members is less than or equal to 1 (2). The statistical tests of differences in proportions are based on OLS regressions. Heteroskedasticity-consistent standard errors (with clustering at the group level) are shown in parentheses.
  * Significant at the 10 percent level (two-sided tests).
  ** Significant at the 5 percent level (two-sided tests).
  *** Significant at the 1 percent level (two-sided tests).

the proportion of group-round observations in the final two rounds in which the mean choice of the three group members is less than or equal to 1 (col. 2) or 2 (col. 3). Own-matched high-ability groups are close to equilibrium statistically significantly more frequently than cross-matched groups and own-matched low-ability groups.

In Section IV.C.1 we report how well our structural level k model of learning can match these differences in convergence according to group composition. In particular, we compare the observed proportions of equilibrium and close-to-equilibrium play to those arising from simulated choices constructed using the estimated parameters of the structural model, and we find that the structural model fits well the broad patterns that we find here.

IV.   Structural Analysis

Section III presented a detailed reduced-form description of how behavior and earnings vary with cognitive ability. To help gain insight into the behavioral mechanisms that underlie these differences, we turn to a

level *k* model of learning. Below we describe our empirical model's specification, our estimation strategy, and our results.

A.   *Level* k *Model of Learning*

We estimate a structural level *k* mixture-of-types model of learning using maximum likelihood. Choices in the first round serve as the initial conditions. Our level *k* model of learning includes nine learner types; and since our model includes rule learners, we distinguish level *k types* from level *k choice rules* (described below). Five fixed level *k* types with $k \in \{0, 1, 2, 3, 4\}$ follow the level *k* choice rule in all rounds $r \geq 2$ (therefore, an individual drawn as a level *k* type follows the same choice rule in every round). We also include four rule learner types who switch from following the level *k* choice rule in round $r = 2$, with $k \in \{0, 1, 2, 3\}$, to following the level $(k + 1)$ choice rule in round $r = 10$. For such L*k*–$(k + 1)$ rule learners, the probability of choosing the level $(k + 1)$ choice rule increases linearly over rounds from 0 in round $r = 2$ to 1 in round $r = 10$ (and is thus given by $[r - 2]/8$ in round r), while the complementary probability of choosing the level *k* choice rule falls linearly over rounds.[17]

Some further notation is necessary to describe the level *k* choice rules. Subject $i \in \{1, 2, 3\}$ in group $g \in \{1, 2, ..., G\}$ is denoted by $ig \in \{1g, 2g, 3g\}$. The choice of subject ig in round $r \in \{1, 2, ..., 10\}$ is denoted by $x_{ig,r} \in \{0, 1, ..., 100\}$. The set of choices in round r of the three subjects in group g is denoted by $\mathbf{x}_{g,r} \equiv \{x_{1g,r}, x_{2g,r}, x_{3g,r}\}$, with mean choice $\overline{x}_{g,r} \equiv (\Sigma_{i=1}^{3} x_{ig,r})/3$.

---

[17] Although the probability of choosing the level $(k + 1)$ choice rule goes up over time, an L*k*–$(k + 1)$ rule learner is allowed to switch back and forth between the level *k* and level $(k + 1)$ choice rules. An alternative specification (AS1) in which rule learners make a once-and-for-all transition to the level $(k + 1)$ choice rule and so cannot switch back fits the data significantly less well (see App. table A3). Note also that in our mixture model framework, we do not model explicitly how rules are chosen; in contrast, Stahl (1996) uses an attraction framework in which a given rule is more attractive the higher its past payoff. We have successfully estimated richer and more parsimonious specifications of our level *k* model of learning. A Bayesian information criterion (BIC)–based model selection exercise selects the preferred specification over a richer model that additionally includes a fixed L5 type (that follows the L5 choice rule in every round $r \geq 2$) and an L4–5 rule learner type. On the basis of the same model selection criterion, the preferred specification also outperforms a more parsimonious model that excludes the fixed L4 type and the L3–4 rule learner type. Additionally, we estimated a specification that extends the preferred specification by including three two-step rule learner types who switch from following the level *k* choice rule in round $r = 2$, with $k \in \{0, 1, 2\}$, to following the level $(k + 2)$ choice rule in round $r = 10$. The BIC values for the preferred specification and this extended specification are the same to three decimal places, and the estimates of the average level *k* choice rules followed by own-matched and cross-matched high– and low–cognitive ability subjects differ little between the two specifications (see table 4 for the preferred specification estimates). However, because of the large number of similar types, we encountered numerical difficulties when trying to obtain standard errors for this extended specification; two-step rule learners are therefore excluded from our analysis.

To model the noise in the choice process, we assume that subjects' choices are independent draws (over rounds and subjects) from discretized and truncated t distributions.[18] Letting $f(\cdot; \mu, \sigma, \nu)$ be the density of the three-parameter t distribution with mean μ, scale σ, and degrees of freedom ν, the probability of a particular choice x by subject ig when following the level k choice rule in round $r \geq 2$ given the group-specific choices in the previous round $\mathbf{x}_{g,r-1}$ is

$$\Pr(x \mid k, \mathbf{x}_{g,r-1}) = [1 - \gamma(r)] \frac{f(x; \mu(k, \mathbf{x}_{g,r-1}), \sigma(\mu, r), \nu)}{\sum_{j=0}^{99} f(j; \mu(k, \mathbf{x}_{g,r-1}), \sigma(\mu, r), \nu)}$$
$$\times \mathbf{1}_{x \in \{0,1,\ldots,99\}} + \gamma(r)\mathbf{1}_{x=100}, \tag{1}$$

where

$$\mu(k, \mathbf{x}_{g,r-1}) = \left(\frac{7}{10}\right)^k \overline{x}_{g,r-1}, \tag{2}$$

rounded to the nearest integer, and $\gamma(r)$ is the probability of choosing 100 in round r given by the empirical frequency of subjects choosing 100 that we observe in that round in our sample (independently of cognitive ability).[19] Therefore, as in Nagel (1995), Stahl (1996), and Duffy and Nagel (1997), a subject following the level 0 choice rule "follows the crowd" in the sense that she aims to copy the average group behavior from the previous round, while a subject who follows the level k choice rule for $k > 0$ best responds to level $(k-1)$ choices in the sense that she aims to hit the current round's target in her group $t_{g,r} \equiv (7/10)\overline{x}_{g,r}$ that would result from everybody in the group noiselessly following the level $(k-1)$ choice rule.[20]

For all level k choice rules the scale parameter σ depends on μ and on the round r in the following way:

$$\sigma(\mu, r) = \exp\left\{\alpha + \left[\mathbf{1}_{\mu \in \{0\}}, \mathbf{1}_{\mu \in \{1\}}, \mathbf{1}_{\mu \in \{2,3,4,5\}}\right]\boldsymbol{\beta} - \delta\left(\frac{r-2}{8}\right)\right\}. \tag{3}$$

[18] Stahl (1996) and Offerman, Potters, and Sonnemans (2002) use truncated normal distributions to model noise in the choice process in structural models of learning. In App. table A3, we show that an alternative specification (AS2) in which the normal distribution replaces the t distribution fits the data significantly less well.

[19] We observe 106 instances of subjects choosing 100 in our sample, making up 1.4 percent of observations. Although such extreme choices are a standard feature in beauty contest data sets, their attraction cannot be explained readily by level k choice rules. Ho et al. (1998, 960) argue that such choices "are probably due to frustration or to misguided attempts to win by single-handedly raising the mean dramatically." In App. table A3, we show that an alternative specification (AS3) in which the choices of 100 come from the same t distribution as for the other choices fits the data significantly less well.

[20] If subjects took into account the effect of their own choice on the target, they would choose lower fractions of $\overline{x}_{g,r-1}$; however, in App. table A3, we show that an alternative specification (AS4) in which $\mu(k, \mathbf{x}_{g,r-1})$ is given by this lower fraction fits the data significantly less well.

In equation (3), $\mathbf{1}_{\mu \in M}$ is an indicator function taking the value one when $\mu \in M$ and taking the value zero otherwise, and $\boldsymbol{\beta}$ is a three-element vector. Thus we allow the variance of the discretized choice distribution to vary with μ in a flexible way and to include a round trend.[21] The number of degrees of freedom ν is common to all level k choice rules.

B. Estimation Strategy

Using the full sample of 780 subjects observed for 10 rounds, we estimate the parameter vector $\boldsymbol{\theta}$ that contains 38 elements. In more detail, we estimate the parameters that describe the probabilities of being the nine different learner types for own-matched high–cognitive ability subjects, cross-matched high-ability subjects, cross-matched low-ability subjects, and own-matched low-ability subjects, giving $8 \times 4 = 32$ parameters. In addition, we estimate the six parameters of the t distribution (α, the three-element vector $\boldsymbol{\beta}$, δ, and ν). Recall that our model includes fixed level k types who follow the same choice rule in every round and rule learners who switch choice rules over time. Letting $k_{ig,r}$ be the level k choice rule that subject ig follows in round r, the set of level k choice rules followed by the three subjects in group g in round r is denoted by $\mathbf{k}_{g,r} \equiv \{ k_{1g,r}, k_{2g,r}, k_{3g,r} \}$, and the set of group g's choice rules for every round $r \geq 2$ is denoted by $\mathbf{k}_g \equiv \{ \mathbf{k}_{g,2}, \mathbf{k}_{g,3}, ..., \mathbf{k}_{g,10} \}$. The 32 parameters that measure the probabilities of the different learner types determine the probability of each different combination of choice rules, so $\mathrm{Pr}(\mathbf{k}_g)$ depends on $\boldsymbol{\theta}$. Finally, letting $\mathbf{x}_g \equiv \{ \mathbf{x}_{g,2}, \mathbf{x}_{g,3}, ..., \mathbf{x}_{g,10} \}$ be the set of group g's choices for every round $r \geq 2$,

$$
\begin{aligned}
\mathrm{Pr}(\mathbf{x}_g \mid \mathbf{k}_g, \mathbf{x}_{g,1}) &= \prod_{r=2}^{10} \mathrm{Pr}(\mathbf{x}_{g,r} \mid \mathbf{k}_{g,r}, \mathbf{x}_{g,r-1}) \\
&= \prod_{r=2}^{10} \prod_{i=1}^{3} \mathrm{Pr}(x_{ig,r} \mid k_{ig,r}, \mathbf{x}_{g,r-1}),
\end{aligned}
\tag{4}
$$

and the likelihood for group g,

$$
L_g(\boldsymbol{\theta} \mid \mathbf{x}_g, \mathbf{x}_{g,1}) = \mathrm{Pr}(\mathbf{x}_g \mid \mathbf{x}_{g,1}) = \sum_{\mathbf{k}_g} \mathrm{Pr}(\mathbf{k}_g) \mathrm{Pr}(\mathbf{x}_g \mid \mathbf{k}_g, \mathbf{x}_{g,1}). \tag{5}
$$

[21] We need the variance to depend on μ and r in order to fit the degree of convergence toward equilibrium play that we observe in the data. The time trend is a function of $r - 2$ since first-round choices serve as the initial conditions and the second round acts as the omitted category. Stahl (1996) also allows the variance of the choice distribution to depend on μ; Ho et al. (1998) allow a time trend in variances.

The sample likelihood is then the product over the G groups of the group likelihoods.

We maximized the sample log likelihood function using a Hessian-based optimization routine. Following Berndt et al. (1974), the Hessian employed in the optimization process was approximated as the sum of outer products of the gradients of the group log likelihoods. The gradients of the group log likelihoods, in turn, were obtained via numerical differentiation. Standard errors were obtained from a Hessian matrix computed using numerical differentiation. We found the optimization problem to be well behaved. In particular, the optimization routine converged to the same parameter vector for multiple sets of starting values, and the Hessian matrix used to obtain standard errors was never found to be close to singular.[22]

C.   Results

In Section IV.C.1, we start by reporting the estimated proportions of learner types arising from our level k model of learning, and we provide evidence that our preferred specification fits well the differences in behavior by cognitive ability observed in the data. In Section IV.C.2, we then show how the subjects' average level k choice rule varies with their own cognitive ability and with that of their opponents. Finally, in Section IV.C.3, we simulate the earnings that accrue to different learner types in order to discover which learner types earn the most.

1.   Estimated Learner Types and Model Goodness of Fit

Table 2 reports the estimated proportions of learner types.[23] In Section IV.C.2 below we summarize this information in terms of average level k choice rules and analyze how the averages vary with own cognitive ability and that of opponents. Instead, our focus here is on the implications of these estimates for behavior in order to see how well our model fits the differences in behavior by cognitive ability in the observed data.

---

[22] According to the model outlined in Sec. IV.A, choice probabilities conditional on the group playing Nash equilibrium in the previous round are identical across learner types. However, there are enough group-round observations in which Nash equilibrium was not played to identify all the learner type probabilities.

[23] The estimated parameters of the t distribution, which complete the estimate of the parameter vector θ, are $\alpha = -2.734$ (0.047); $\beta = [-2.778$ (0.102), -1.826 (0.079), -1.134 (0.066)]; $\delta = 0.844$ (0.088); and $\nu = 0.749$ (0.035). Standard errors are in parentheses. As reported in table A3, the maximized log likelihood is $-21,510$.

TABLE 2
ESTIMATED LEARNER TYPE PROBABILITIES

	OWN-MATCHED SUBJECTS		CROSS-MATCHED SUBJECTS	
	High Ability (1)	Low Ability (2)	High Ability (3)	Low Ability (4)
L0 type	.002	.023*	.004	.021
	(.009)	(.016)	(.009)	(.018)
L1 type	.222***	.280***	.172***	.409***
	(.045)	(.050)	(.052)	(.065)
L2 type	.148***	.034	.174***	.281***
	(.053)	(.047)	(.067)	(.069)
L3 type	.000	.000	.053	.000
	. . .	. . .	(.051)	. . .
L4 type	.001	.000	.000	.013
	(.036)	. . .	. . .	(.026)
L0–1 rule learner type	.043**	.113***	.052**	.090**
	(.025)	(.034)	(.029)	(.043)
L1–2 rule learner type	.218***	.256***	.334***	.061
	(.055)	(.057)	(.069)	(.065)
L2–3 rule learner type	.177***	.188***	.186***	.078*
	(.055)	(.051)	(.071)	(.061)
L3–4 rule learner type	.189***	.106***	.026	.047
	(.058)	(.029)	(.032)	(.044)
Proportion of rule learners	.627***	.663***	.597***	.275***
	(.072)	(.070)	(.097)	(.092)

NOTE.—Standard errors are shown in parentheses.
* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.

To give a visual impression of how well the model fits the observed data, we simulate choices over the 10 rounds using the estimated parameters. Figure 7 shows the simulated and observed paths of average behavior for own- and cross-matched subjects: we see that the observed and simulated paths match closely (the notes to the figure explain how the simulated paths were constructed).

Figure 8A shows the simulated path of earnings for cross-matched subjects (recall that, by construction, own-matched high-ability subjects and own-matched low-ability subjects must earn $2.00 on average in every round). We fit the pattern of divergence in earnings over rounds, although the magnitude of the divergence is not quite as big as that observed in the data (see fig. 4B in Sec. III.B.3). Our simulations return a divergence in earnings over rounds even though, as in the observed data, high- and low-ability cross-matched subjects' simulated choices remain similar throughout the experiment. Rule learning is crucial to explaining the pattern of divergence: table 2 shows that there are 60 percent of rule learners among cross-matched high-ability subjects, but only 28 percent among cross-matched low-ability subjects (the difference is statistically significant
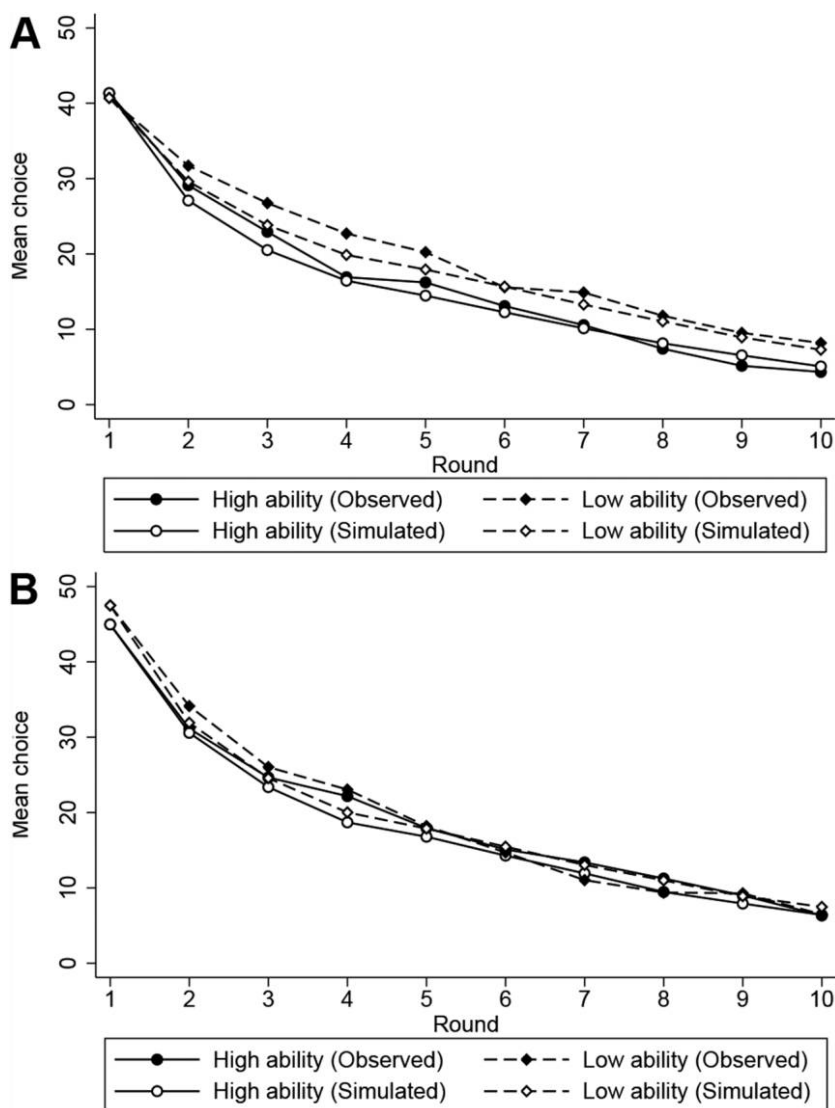
FIG. 7.—Observed and simulated round-by-round means of choices. *A*, Own-matched subjects. *B*, Cross-matched subjects. Simulated choices were obtained using a sequential method. Specifically, for each of the 260 groups in the experimental sample, we drew the type of each group member from the appropriate estimated distribution of learner types reported in table 2. We then simulated the choice of each of the three group members in round 2, given the observed behavior, specifically the group average, in the first round (in the figures, the simulated first-round data are given by the observed data). Next, we simulated the choice of each of the three group members in round 3, given the average of the simulated choices of the group members in round 2. We continued sequentially in this manner to round 10, with simulated choices in each round being based on the group-level average of the simulated choices in the previous round. This procedure was repeated 100 times for each of the 260 groups.
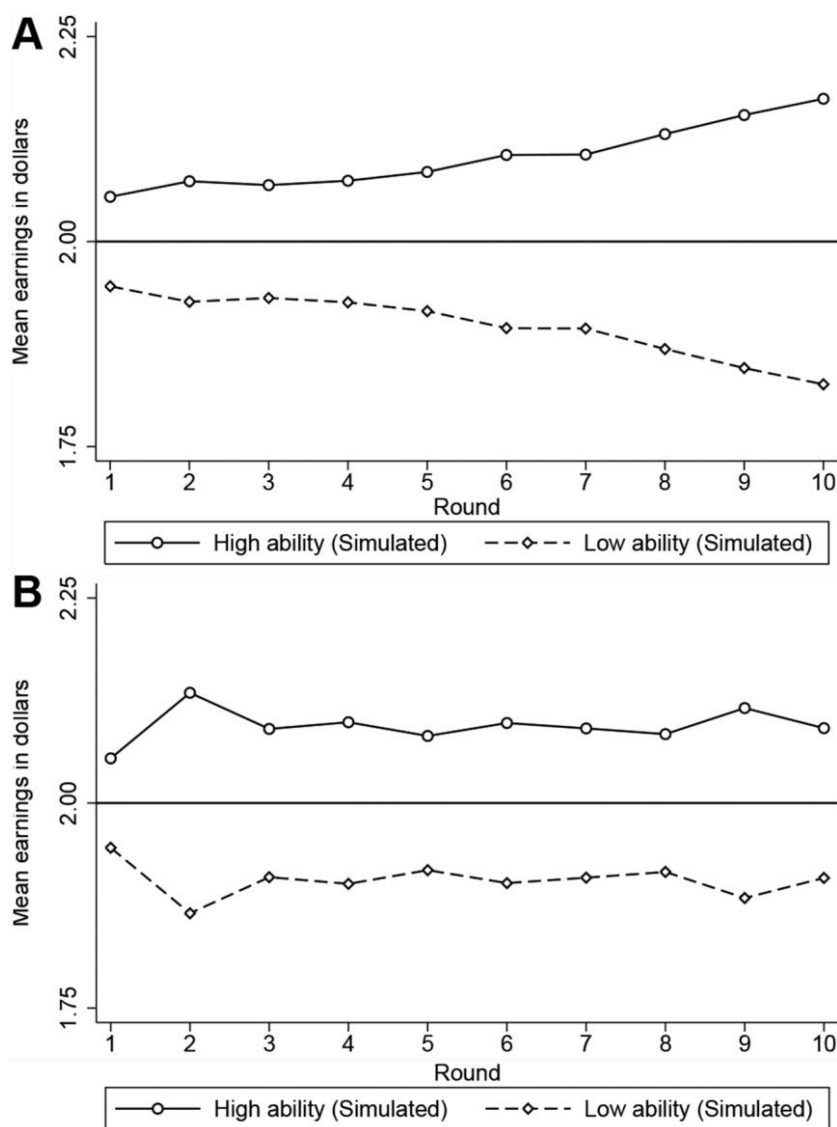
FIG. 8.—Simulated round-by-round means of earnings of cross-matched subjects. *A*, Preferred specification. *B*, Without rule learners. Simulated earnings were computed from the simulated choices generated as described in the notes to figure 7.

with a two-sided $p = .016$;[24] and figure 8*B* shows that when we reestimate the model without rule learners, we can no longer fit the divergence in

[24] The difference of 0.322 between the proportion of rule learners among cross-matched high-ability subjects (0.597) and the proportion among cross-matched low-ability subjects (0.275) appears to be a robust feature of our sample. As noted above, the differ-

earnings over rounds. Further evidence that rule learning plays an important role in explaining subjects' choices more generally comes from the estimates of an alternative specification of the structural model without rule learners: Appendix table A3 shows that this alternative specification (AS5) fits the data significantly less well.

Table 3 shows that the simulated choices match quite well the extent of equilibrium and close-to-equilibrium play that we see in the data. In particular, the simulated choices fit well the broad pattern of increasing convergence as the proportion of high-ability subjects in the group goes up.

Section B in the Appendix provides further evidence that our structural level k model of learning fits well. In particular, table A3 reports values of log likelihoods and of sums of squared deviations of choices, earnings, and convergence statistics computed from simulated choices and provides evidence that our preferred specification fits better than various alternative specifications (some nested and others not), while table A4 shows that our model continues to perform well out of sample.

2. The Impact of Cognitive Ability on Average Level k Choice Rules

We now analyze how learner types vary with subjects' own cognitive ability and that of their opponents. To do this, we summarize the estimated proportions of learner types in a single statistic measuring the average level k choice rule followed by the subjects (the notes to table 4 provide the details of how these averages of choice rules are computed). From table 4, we can see that, across all rounds, the average level k choice rule followed by own-matched high-ability subjects is 1.97, while for own-matched low-ability subjects the average level is 1.63, with the difference statistically significant at the 1 percent level. For cross-matched subjects, the difference is less pronounced but still evident: the average level fol-

ence is statistically significant at the 5 percent level. To test how sensitive our estimate of the difference is to our particular sample, we ran 200 Monte Carlo simulations using a data generation process based on our estimated model, except that we assumed that the proportion of rule learners among cross-matched low-ability subjects in the population matches the proportion that we estimate for the cross-matched high-ability subjects (i.e., 0.597). The simulations show that if our subjects were drawn from a population in which there was no difference in the proportion of rule learners, then there would be a less than 1 percent chance of drawing a sample that gives an estimated difference greater than or equal to our actual estimate of 0.322 (and only a 5 percent chance of estimating a difference greater than or equal to 0.200). We have no direct evidence about what might drive the difference in the proportion of rule learners. One conjecture is that the earnings difference reported in Sec. III.B.3 between cross-matched high- and low-ability subjects demotivates the low-ability subjects, thus leading to less rule learning. By contrast, the proportion of rule learners among own-matched subjects is almost the same for high- and low-ability subjects (table 2), and by construction own-matched high-ability subjects earn the same, on average, in every round as own-matched low-ability subjects (Sec. III.A).

TABLE 3
OBSERVED AND SIMULATED PROPORTIONS OF EQUILIBRIUM AND
CLOSE-TO-EQUILIBRIUM PLAY IN ROUNDS 9 AND 10

	Equilibrium (1)	Group Mean ≤1 (2)	Group Mean ≤2 (3)
Own-matched high-ability groups:			
Observed	.293	.453	.540
Simulated	.201	.357	.471
Cross-matched groups:			
Observed	.145	.282	.368
Simulated	.109	.228	.336
Own-matched low-ability groups:			
Observed	.127	.247	.313
Simulated	.107	.218	.317

NOTE.—Simulated proportions were computed from the simulated choices generated as described in the notes to fig. 7.

lowed by cross-matched high-ability subjects is 1.76, while for cross-matched low-ability subjects the average level is 1.52, with the difference significant at the 5 percent level. The average level followed by cross-matched high-ability subjects is higher even though they face a lower number of high-ability opponents on average than do cross-matched low-ability subjects.

Thus subjects' own cognitive ability type has a significant effect on the average level that they follow: high–cognitive ability subjects follow level

TABLE 4
ESTIMATED AVERAGES OF LEVEL k CHOICE RULES

OWN-MATCHED SUBJECTS			CROSS-MATCHED SUBJECTS			DIFFERENCE-IN-DIFFERENCE
High Ability	Low Ability	Difference	High Ability	Low Ability	Difference	
1.974***	1.629***	.346***	1.759***	1.520***	.240**	.106
(.074)	(.070)	(.101)	(.067)	(.077)	(.102)	(.144)

HIGH-ABILITY SUBJECTS			LOW-ABILITY SUBJECTS			DIFFERENCE-IN-DIFFERENCE
Own-Matched	Cross-Matched	Difference	Cross-Matched	Own-Matched	Difference	
1.974***	1.759***	.215**	1.520***	1.629***	−.109	.324**
(.074)	(.067)	(.099)	(.077)	(.070)	(.104)	(.144)

NOTE.—The average choice rule of an Lk type is k, as this type follows the level k choice rule in every round. The average choice rule of an Lk–$(k + 1)$ rule learner is $k + (1/2)$. We use the estimated proportions of learner types from table 2 to compute the average level k choice rule over all learner types. Tests for the significance of differences and of differences-in-differences are two-sided. Standard errors are shown in parentheses.
    * Significant at the 10 percent level.
   ** Significant at the 5 percent level.
  *** Significant at the 1 percent level.

k choice rules that are significantly higher on average than those followed by low-ability subjects. Further evidence that cognitive ability plays an important role in determining subjects' choice rules comes from estimating an alternative specification of the structural model in which the probabilities of the different learner types are not allowed to vary with subjects' own cognitive ability or with whether subjects are in own-matched or cross-matched groups. Appendix table A3 shows that this alternative specification (AS6) fits the data significantly less well.

We also find significant differences in how average level k choice rules respond to the cognitive ability of opponents. We noted in Section II.B that, on average, high-ability subjects face four-thirds more high-ability opponents in own-matched groups than in cross-matched groups, while low-ability subjects face four-thirds more high-ability opponents in cross-matched groups than in own-matched groups. The second row of table 4 shows that high–cognitive ability subjects respond to the cognitive ability of their opponents: across all rounds, the average level followed by own-matched high–cognitive ability subjects is 0.21 higher than that followed by cross-matched high-ability subjects, with the difference significant at the 5 percent level. Low–cognitive ability subjects, on the other hand, do not respond to the cognitive ability of their opponents: the average level followed by cross-matched low-ability subjects is not statistically significantly different from that followed by own-matched low-ability subjects (two-sided $p = .295$). Furthermore, table 4 shows that the difference-in-differences in how the average levels followed by high– and low–cognitive ability subjects respond to the cognitive ability of opponents is statistically significant (a two-sided test shows significance at the 5 percent level).

Thus subjects' own cognitive ability has a significant effect on how the average level that they follow responds to the cognitive ability of their opponents: the average level k choice rule followed by high–cognitive ability subjects responds significantly to the number of high-ability opponents that they face, while the average level followed by low–cognitive ability subjects does not respond. Further evidence that cognitive ability plays a role in determining how subjects' choice rules respond to the cognitive ability of opponents comes from estimating an alternative specification of the structural model in which the probabilities of the different learner types are not allowed to vary with whether subjects are in own-matched or cross-matched groups (but are allowed to vary with subjects' own cognitive ability). Appendix table A3 provides evidence that this alternative specification (AS7) fits the data less well.[25]

---

[25] In our experiment, subjects can condition current behavior on (i) information about the cognitive ability type composition of the group and (ii) the previous behavior of the group members. Our design therefore allows us to identify how group composition affects current behavior conditional on past group behavior. Our design does not allow us to identify how subjects would respond to a given history if they did not know the group composition.

3.   Levels and Earnings

Finally, we turn to the question of which learner types earn the most. The estimates of the structural model's parameters also allow us to simulate the earnings of each learner type, according to subjects' own cognitive ability and that of their opponents. Table 5 shows how simulated total earnings over rounds 2–10 vary according to learner type (the notes to the table describe how the simulated earnings were computed). The differences in earnings by learner type are large: for instance, L2 types earn about 40 percent more than L1 types across the board. In every case, the three best-performing learner types are L3 types, L4 types, and L3–4 rule learners, although the order varies. This means that it is always optimal to be a high-level learner type, even for subjects facing a large proportion of low-ability opponents who tend to choose higher numbers. In order of decreasing earnings, the order among the other six types is always L2–3 rule learners, L2 types, L1–2 rule learners, L1 types, L1–0 rule learners, and finally, L0 types.

Given the estimated proportions of learner types reported in table 2 and the average level k choice rules that subjects follow reported in table 4, it is clear that subjects are constrained in their levels below those that are optimal, given the learner types of their opponents. Of course, unless subjects all play equilibrium, their level k choice rules will tend to be too low on average; nonetheless, subjects are leaving a substantial amount of money on the table.

On average, own-matched high-ability subjects earn $18.00 across rounds 2–10 and so are leaving $4.59 on the table compared to the pay-

TABLE 5
SIMULATED TOTAL EARNINGS OVER ROUNDS 2–10 (in Dollars) BY LEARNER TYPE

	OWN-MATCHED SUBJECTS		CROSS-MATCHED SUBJECTS	
	High Ability (1)	Low Ability (2)	High Ability (3)	Low Ability (4)
L0 type	7.83	8.41	8.69	8.25
L1 type	14.00	15.22	14.75	14.22
L2 type	19.14	21.03	21.22	20.70
L3 type	21.96	23.53	24.27	23.85
L4 type	22.59[a]	23.18	23.42	24.03[a]
L0–1 rule learner type	10.35	11.49	11.96	10.89
L1–2 rule learner type	16.22	18.44	18.42	17.35
L2–3 rule learner type	21.35	22.31	22.90	21.68
L3–4 rule learner type	22.50	23.75[a]	24.33[a]	23.86

NOTE.—The earnings of a player of a particular learner type were computed on the basis of simulated choices constructed using a method similar to that described in the notes to fig. 7 in which opponents' types were drawn from the appropriate estimated distribution of learner types from table 2.

[a] The learner type that maximizes earnings according to subjects' own cognitive ability and that of their opponents.

off-maximizing L4 type (recall from table 4 that the average level k choice rule followed by these subjects is only 1.97). Own-matched low-ability subjects also earn $18.00 on average and so are leaving $5.75 on the table compared to the payoff-maximizing L3–4 rule learner type (from table 4, their average level is 1.63). Thus, even though high- and low-ability own-matched subjects earn the same amount on average by construction, low-ability subjects leave more money on the table.

Cross-matched high-ability subjects earn $19.73 across rounds 2–10 and so are leaving $4.61 on the table compared to the payoff-maximizing L3–4 rule learner type (their average level is 1.76). Cross-matched low-ability subjects earn $16.27 on average and so are leaving $7.75 on the table compared to the payoff-maximizing L4 type (their average level is just 1.52). The L1 types are the most common among cross-matched low-ability subjects, making up 41 percent of the total: they earn a striking $9.80 less on average than the payoff-maximizing learner type.

Finally, by not adjusting their level upward on average, the low–cognitive ability subjects lose out relative to the payoff-maximizing type when the cognitive ability of their opponents goes up: the low-ability subjects leave $2.01 more on the table in cross-matched groups than in own-matched low-ability groups (from above, $7.75 minus $5.75). In contrast, the high-ability subjects adjust their level upward and do not lose out: the high-ability subjects leave the same amount of money on the table in own-matched high-ability groups and in cross-matched groups (from above, $4.59 and $4.61).

V.   Comparing the Effects of Cognitive Ability and Character Skills

As well as exploring how cognitive ability influences behavior, success, and the evolution of play toward Nash equilibrium in repeated strategic interactions, we are also interested in discovering the effects of character skills. To this end, we administered questionnaires to measure character skills of a subset of our subjects. Section V.A describes our measures of character skills, Section V.B presents a reduced-form analysis comparing the behavioral effects of cognitive ability and character skills, while Section V.C builds on our structural level k model of learning to compare the influence of cognitive ability and character skills on the level k choice rules followed by our subjects.

A.   Measuring Character Skills

As noted in Section II.A, a subset of our sessions included questionnaires to measure character skills that were administered before the test of cognitive ability. We ran these 15 sessions in January 2014, with 18 sub-

jects in each session and 270 participants in total. Apart from administering the questionnaires, these sessions were run in the manner described in Section II. All the sessions were own-matched, giving 45 groups of three own-matched high–cognitive ability subjects and 45 groups of three own-matched low–cognitive ability subjects.

We measured seven character skills: openness, conscientiousness, extraversion, agreeableness, emotional stability (the Big Five), grit, and consideration of future of consequences (CFC). We measured the Big Five using the 44-item Big Five Inventory (John et al. 1991, 2008), grit using the 12-item Grit Scale (Duckworth et al. 2007), and CFC using the 12-item CFC Scale (Strathman et al. 1994), in that order.[26] These three questionnaires use a five-point Likert scale to measure responses to individual questions.[27] We gave the subjects 8 minutes to complete the 68 questions, which were displayed on 17 screens.[28] The subjects could move back and forth between the 17 screens, and time permitting, they were allowed to change their previous answers. We did not provide any monetary incentives for answering the questions. The experimental instructions provided are in the supplementary web appendix.

Table 6 presents the correlation matrix for our seven measures of character skills. Many of the correlations are high. For instance, we find a correlation of .63 between conscientiousness and grit, of .39 between grit and CFC, of .38 between conscientiousness and CFC, and of .34 between agreeableness and emotional stability. A further five correlations are above .2.[29]

---

[26] John et al. (2008) provide a detailed description and history of the Big Five. Borghans et al. (2008, 1008) argue that it is more helpful to use the term "emotional stability" rather than its obverse, "neuroticism." Grit is defined as "perseverance and passion for long-term goals" (Duckworth et al. 2007, 1087) and is viewed as a measure of persistence (Borghans et al. 2011, 318). CFC is a survey-based measure of time preference: in particular, CFC is designed to measure "a stable individual difference in the extent to which people consider distant versus immediate consequences of potential behaviors" (Strathman et al. 1994, 742). Houser and Winter (2004) and Daly, Harmon, and Delaney (2009) find a statistically significant correlation between CFC and discount rates measured using incentivized choice tasks. Recent papers have found a relationship between CFC and behaviors such as savings (Webley and Nyhus 2013), study habits (Delaney, Harmon, and Ryan 2013), and willingness to undertake training (Fouarge, Schils, and de Grip 2013), while Strathman et al. (1994) find a relationship with concern for health, alcohol consumption, smoking, and environmentally friendly behavior.

[27] Three-tenths of a percent of the responses are missing (57 of $270 \times 68 = 18{,}360$). For each question, we replaced any missing responses by the sample average of the nonmissing responses to that particular question.

[28] The Big Five Inventory alone is designed to take about 5 minutes (John et al. 2008, 137).

[29] The correlations between grit and the Big Five are similar to those found by Duckworth et al. (2007): $r = .77$ (study 2) and $r = .64$ (study 5) for conscientiousness; $r = .38$ for emotional stability; $r = .24$ for agreeableness; $r = .22$ for extraversion; and $r = .14$ for openness. Strathman et al. (1994) find a correlation of .49 between CFC and conscientiousness (they did not measure the other Big Five traits).

TABLE 6
CORRELATIONS BETWEEN MEASURES OF CHARACTER SKILLS

Character Skill	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability	Grit	CFC
Openness	1.000						
Conscientiousness	−.045	1.000					
Extraversion	.130	.111	1.000				
Agreeableness	.097	.211	.112	1.000			
Emotional stability	.024	.230	.181	.341	1.000		
Grit	−.041	.629	.242	.183	.275	1.000	
CFC	.246	.381	.174	.041	.085	.388	1.000

TABLE 7
FACTOR LOADINGS

Character Skill	Factor 1 (1)	Factor 2 (2)	Factor 3 (3)
Openness	−.085	.043	.886
Conscientiousness	.845	.156	−.108
Extraversion	.221	.295	.440
Agreeableness	.067	.797	.070
Emotional stability	.199	.770	−.009
Grit	.850	.195	−.041
CFC	.668	−.162	.483

The high degree of correlation between our seven measures of character skills justifies the construction of a smaller number of uncorrelated factors to identify the effects of character skills on behavior and compare the effects to that of cognitive ability. We undertake a principal component factor analysis with Varimax rotation, implemented to give factors that are uncorrelated with each other (see Jolliffe 1995). Before rotation, three factors have eigenvalues above one, and so these are retained in the rotation.[30] These three factors explain 66 percent of the variance (factor 1 explains 28 percent, factor 2 explains a further 20 percent, and factor 3 explains the final 18 percent). The factors have zero mean and unit variance by construction.

Table 7 presents the factor loadings. As we can see from the table, factor 1 mainly captures conscientiousness, grit, and CFC; factor 2 mainly captures agreeableness and emotional stability; while factor 3 mainly captures openness, extraversion, and CFC. The correlations between cognitive ability and these factors are low and are reported in the fifth paragraph of Section V.B.2.

B. Reduced-Form Comparison of Cognitive Ability and Character Skills

In this section, we compare the behavioral effects of cognitive ability and character skills. In Section V.B.1, we first describe the influence of character skills on round-by-round choices. In Section V.B.2, we compare the effects of cognitive ability and character skills on average choices and earnings. Finally, in Section V.B.3, we examine the influence of cognitive ability and character skills on learning at the group level.

---

[30] Retaining factors with eigenvalues above one is a standard criterion for choosing the number of factors due to Kaiser (1960).

1. Round-by-Round Choices according to Character Skills

We start by providing a visual description of how the three factors capturing character skills influence behavior in the p beauty contest. Figure 9 shows the round-by-round evolution of mean choices for subjects categorized as high (above the sample median) or low (below the sam-
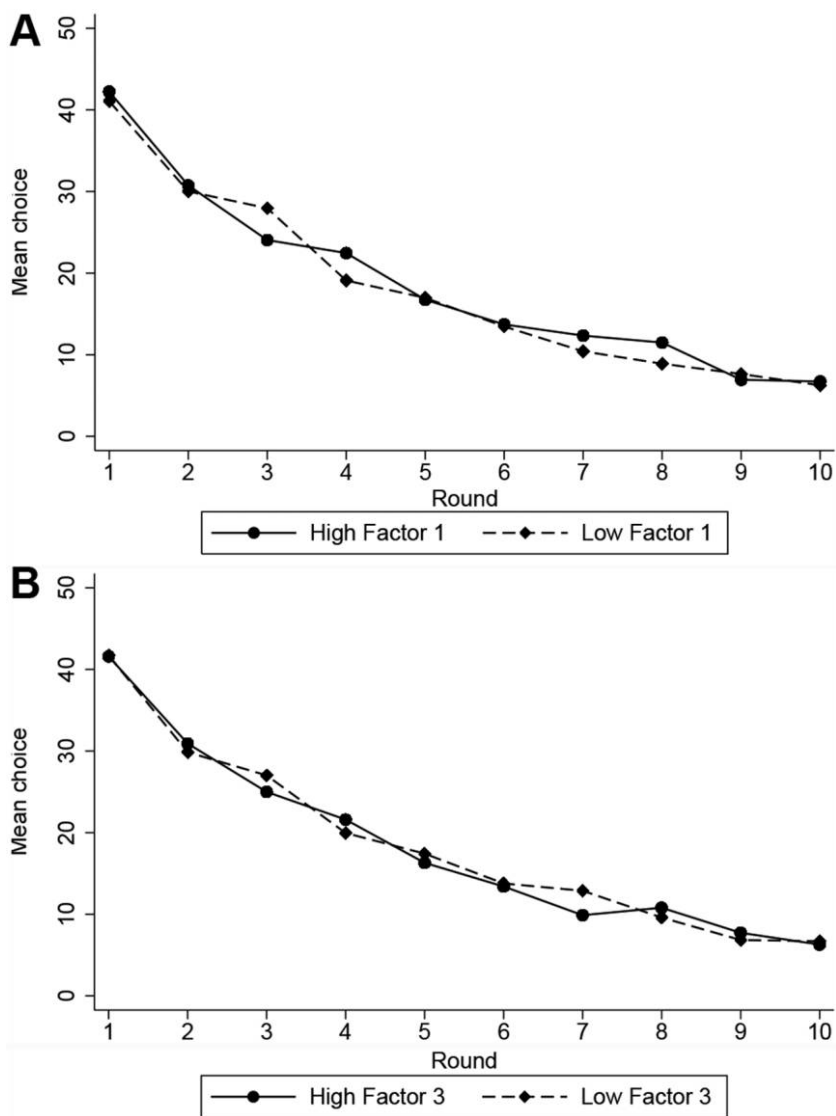

FIG. 9.—Round-by-round means of p beauty contest choices: factor 1 and factor 3

ple median) on factor 1 and factor 3. Across all 10 rounds, the mean choice of high–factor 1 subjects is not statistically significantly different from that of low–factor 1 subjects (two-sided $p = .608$), and the same is true for factor 3 (two-sided $p = .862$).

However, as illustrated in figure 10, across the 10 rounds the mean choice of subjects high on factor 2 is 2.7 lower than that of subjects low on factor 2, and the difference is highly statistically significant (two-sided $p = .010$). By the final round, the mean choice of high–factor 2 subjects falls to 4.9, while that of low–factor 2 subjects falls to 8.1, and the difference of 3.3 is also statistically significant (two-sided $p = .033$). Thus, subjects high on factor 2, who are more agreeable and emotionally stable, tend to choose numbers closer to equilibrium, both on average across the 10 rounds and in the final round after a period of learning.

The statistical tests reported above are based on OLS regressions and use heteroskedasticity-consistent standard errors. We cluster at the group level to allow for within-group nonindependence across rounds. The regressions use the 270 subjects whose character skills we measured, observed for 10 rounds or, in the last case, just in round 10.

2. Comparison of the Effects on Average Choices and Earnings

Next, we compare the effects of cognitive ability and character skills on choices and earnings across the 10 rounds. We run linear regressions
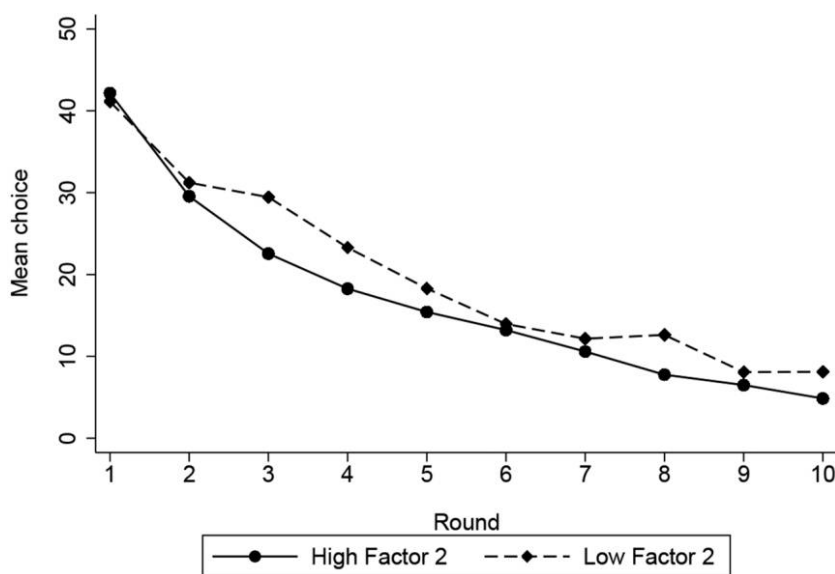

Fig. 10.—Round-by-round means of p beauty contest choices: factor 2

of choices and earnings on subjects' score in the Raven test and on the three factors that capture character skills. Recall that we matched subjects by cognitive ability type but not according to their character skills. Thus, these regressions measure the effects of cognitive ability and character skills within matching group; that is, they measure the effects of cognitive ability and character skills within the set of subjects who were matched into groups of three high–cognitive ability subjects and within the set of subjects who were matched into groups of three low–cognitive ability subjects. Since the factors have zero mean and unit variance by construction, we standardize the Raven test scores to also have zero mean and unit variance.

Table 8 reports the results of the regressions of average choices on cognitive ability and character skills. Panel A reports the results of regressions that use all 270 subjects whose character skills we measured, while panels B and C report the results of regressions that use only the 135 own-matched high–cognitive ability subjects (panel B) and the 135 own-matched low-ability subjects (panel C). In each case, model 1 includes only cognitive ability as an explanatory variable, model 2 includes only character skills, and model 3 includes both cognitive ability and character skills.

Looking first at model 3 in panel A of table 8, we see that both cognitive ability and character skills influence behavior. In particular, a one standard deviation increase in the Raven test score reduces subjects' mean choice across the 10 rounds by 2.2, while a one standard deviation increase in factor 2, which captures an increase in agreeableness and emotional stability, reduces choices by 1.2. These reductions represent decreases of 12 percent and 7 percent, respectively, in choices relative to the mean. Factors 1 and 3 do not have a statistically significant effect on behavior. Thus, the impact of cognitive ability on behavior is about 75 percent bigger than that of character skills, which operates through factor 2.

Second, looking at model 3 in panels B and C of table 8, we see that the effect of increasing cognitive ability on behavior operates mainly through the subjects who are already high in cognitive ability, while character skills affect the behavior of both high– and low–cognitive ability subjects. Within the subjects already high in cognitive ability, a one standard deviation increase in the Raven test score reduces subjects' mean choice by 4.0, a reduction of 22 percent relative to the mean.[31]

---

[31] Note that we can directly compare the parameter estimates in panels B and C of table 8 since the Raven test scores and factors were standardized before splitting the sample by cognitive ability type. As a result, a single regression with dummy variables for high– and low–cognitive ability subjects together with interactions of those dummies with Raven test scores and the three factors returns estimates identical to those reported in model 3 of panels B and C of table 8. The same is true for model 1 when the factors are excluded from the single regression and for model 2 when the standardized Raven test score is excluded from the single regression.

TABLE 8
Effects of Cognitive Ability and Character Skills
on Mean p Beauty Contest Choices

	(1)	(2)	(3)
	A. All Subjects		
Standardized Raven test score	−2.142**		−2.177**
	(.971)		(.985)
Factor 1		.360	.153
		(.517)	(.530)
Factor 2		−1.282**	−1.238**
		(.508)	(.484)
Factor 3		.384	.411
		(.604)	(.585)
Subject-round observations	2,700	2,700	2,700
	B. Own-Matched High–Cognitive Ability Subjects		
Standardized Raven test score	−3.935***		−4.029***
	(1.018)		(1.024)
Factor 1		.111	−.347
		(.475)	(.522)
Factor 2		−1.049*	−1.082*
		(.618)	(.587)
Factor 3		.219	.309
		(.767)	(.767)
Subject-round observations	1,350	1,350	1,350
	C. Own-Matched Low–Cognitive Ability Subjects		
Standardized Raven test score	−.556		−.636
	(1.555)		(1.541)
Factor 1		.522	.509
		(.905)	(.899)
Factor 2		−1.353*	−1.372*
		(.740)	(.747)
Factor 3		.552	.559
		(.899)	(.886)
Subject-round observations	1,350	1,350	1,350

Note.—The models are linear and are estimated using OLS regressions. The regressions in panel A use the 270 subjects whose character skills we measured, observed for 10 rounds. For all regressions, we standardize the Raven test scores using these 270 subjects. All models include an intercept. In panel A, models 1 and 3 include a dummy for whether subjects are own-matched high–cognitive ability or own-matched low–cognitive ability (the coefficients on the dummy are not statistically significantly different from zero). Heteroskedasticity-consistent standard errors with clustering at the group level are shown in parentheses.

* Significant at the 10 percent level (two-sided tests).
** Significant at the 5 percent level (two-sided tests).
*** Significant at the 1 percent level (two-sided tests).

Third, comparing the estimates across models 1–3, we see that the estimated influence of cognitive ability on behavior does not depend on whether we control for character skills, and vice versa. Consistent with this, our measures of cognitive ability and character skills are not strongly correlated: we find a correlation of −.14 between cognitive ability and fac-

tor 1, of .07 between cognitive ability and factor 2, and of .01 between cognitive ability and factor 3.[32]

In summary, table 8 shows that (i) both cognitive ability and character skills influence behavior in our repeated strategic interaction; (ii) subjects higher in cognitive ability and higher in factor 2 (which captures agreeableness and emotional stability) choose numbers closer to equilibrium; (iii) changes in cognitive ability have a more powerful effect on behavior than changes in character skills; (iv) marginal changes in cognitive ability mainly affect the behavior of subjects who are already cognitively able, while changes in character skills influence the behavior of subjects who are high or low in cognitive ability; and (v) the estimated influence of cognitive ability does not depend on whether we control for character skills, and vice versa.

Table 9 shows that the effects of cognitive ability and character skills on behavior translate into differences in success in the beauty contest, in terms of earnings per round, the log of earnings per round, and the per-round probability of winning the contest (where "winning" is defined to include ties). A one standard deviation increase in the Raven test score increases earnings per round by $0.11, that is, an increase of 6 percent relative to the mean, while a one standard deviation increase in factor 2 increases earnings per round by $0.13. Thus the effects of cognitive ability and character skills on earnings have a similar magnitude in a linear model. Consistent with the absence of an effect on choices, factors 1 and 3 do not have a statistically significant effect on earnings or the probability of winning.

In Section E of the Appendix, we dig deeper into the interplay of cognitive ability and character skills and show that our main findings are robust to allowing interactions between cognitive ability and character skills and to including cognitive ability and character skill dummies in our regressions. Furthermore, we find no statistically significant interactions between cognitive ability and character skills, except for a weakly statistically significant negative interaction between cognitive ability and factor 1 in the case of behavior. That is, we find some evidence that marginal changes in cognitive ability reduce mean choices more for subjects with a higher factor 1 score (who are more conscientious and gritty).

3. Learning at the Group Level

Our reduced-form analysis concludes with an examination of the influence of cognitive ability and character skills on learning at the group

---

[32] As noted by Almlund et al. (2011, 46), most character skills are weakly correlated with cognitive ability, particularly so for measures of fluid intelligence (43), but character skills may nonetheless have some influence on measured cognitive ability (sec. 5.6).

TABLE 9
EFFECTS OF COGNITIVE ABILITY AND CHARACTER SKILLS ON
PER-ROUND EARNINGS AND WINNING PROBABILITY

	Earnings ($ per Round) (1)	Log of Earnings (2)	Winning Probability (3)
Standardized Raven test score	.113*	.075**	.040***
	(.067)	(.032)	(.015)
Factor 1	.031	.001	−.006
	(.051)	(.024)	(.011)
Factor 2	.129**	.057**	.023**
	(.051)	(.022)	(.010)
Factor 3	−.048	−.013	−.002
	(.047)	(.021)	(.010)
Subject-round observations	2,700	2,700	2,700

NOTE.—The models are linear and are estimated using OLS regressions. The regressions use the 270 subjects whose character skills we measured, observed for 10 rounds. We standardize the Raven test scores using these 270 subjects. All models include an intercept and a dummy for whether subjects are own-matched high–cognitive ability or own-matched low–cognitive ability. When taking the log of per-round earnings, we add $0.50 to earnings in each round (the show-up fee of $5.00 divided by the number of rounds) to avoid taking the log of zero. "Winning" is defined to include a tie with one or both other group members. The statistical significances of the coefficient estimates in the per-round winning probability model are robust to estimating a probit model. Heteroskedasticity-consistent standard errors with clustering at the group level are shown in parentheses.
* Significant at the 10 percent level (two-sided tests).
** Significant at the 5 percent level (two-sided tests).
*** Significant at the 1 percent level (two-sided tests).

level. In particular, we look at the effect of group-level averages of cognitive ability and character skills on how close our groups of three subjects come to equilibrium play toward the end of the 10 rounds. We analyze only behavior, since average earnings must always be $2.00 per round at the group level.

Table 10 reports the results of regressions of group-level average choices in round 10, rounds 9 and 10, and rounds 8–10 on the group-level averages of cognitive ability and factors 1–3. We find that cognitive ability and character skills influence learning at the group level: groups with higher cognitive ability on average and higher factor 2 on average, which captures higher agreeableness and emotional stability, end up closer to the Nash equilibrium in the last few rounds. For instance, a one-unit increase in a group's average standardized Raven test score lowers the group's average choice by 5.0 in the last three rounds, while a one-unit increase in a group's average factor 2 lowers the group's average choice by 3.8.[33] These reductions represent decreases of 62 percent

[33] Note that the standard deviation of the group-level mean Raven test score is 0.84, while the standard deviation of the group-level mean factor 2 is 0.54. The first standard deviation is higher because subjects are matched by cognitive ability.

TABLE 10
EFFECTS OF GROUP-LEVEL MEAN COGNITIVE ABILITY AND CHARACTER SKILLS ON
GROUP-LEVEL MEAN p BEAUTY CONTEST CHOICES IN THE LAST FEW ROUNDS

	Round 10 (1)	Rounds 9 and 10 (2)	Rounds 8–10 (3)
Group mean of standardized Raven test score	−5.059**	−3.873*	−4.988**
	(2.007)	(2.234)	(2.351)
Group mean of factor 1	.159	.125	.378
	(1.568)	(1.657)	(1.821)
Group mean of factor 2	−5.423***	−4.005***	−3.845***
	(1.871)	(1.268)	(1.376)
Group mean of factor 3	−.744	−.639	.404
	(1.467)	(1.376)	(1.508)
Group-round observations	90	180	270

NOTE.—The models are linear and are estimated using OLS regressions. The regressions use the 270 subjects whose character skills we measured, giving 90 groups observed in round 10, rounds 9 and 10, or rounds 8–10. We standardize the Raven test scores using these 270 subjects, and we calculate group means of the scores standardized at the individual level. All models include an intercept and a dummy for whether groups are own-matched high–cognitive ability or own-matched low–cognitive ability (the coefficients on the dummy are not statistically significantly different from zero). Heteroskedasticity-consistent standard errors with clustering at the group level are shown in parentheses.
  * Significant at the 10 percent level (two-sided tests).
 ** Significant at the 5 percent level (two-sided tests).
*** Significant at the 1 percent level (two-sided tests).

and 48 percent, respectively, in choices relative to the mean in the last three rounds. Once again, factors 1 and 3 do not have statistically significant effects.

C. Level k Comparison of Cognitive Ability and Character Skills

Finally, we compare the effects of cognitive ability and character skills on the adoption of level k choice rules. Similarly to the reduced-form analysis reported in Section V.B.2, we allow the probabilities of adopting the level k choice rules to vary in subjects' score in the Raven test and their scores on the three factors that capture character skills.

To do this, we modify our structural level k model of learning from Section IV.A. As before, this modified model includes five fixed level k learner types with $k \in \{0, 1, 2, 3, 4\}$, who follow the level k choice rule in all rounds $r \geq 2$. We exclude the four rule learner types that were also included in Section IV.A.[34] A subject's probability of being each of the

---

[34] We do this because of empirical difficulties encountered when estimating the effects of cognitive ability and character skills on type probabilities in a model with many distinct types.

five learner types depends on her cognitive ability and character skills as follows:

$$\Pr(k \mid \mathbf{z}_{ig}) = \frac{\exp(\boldsymbol{\lambda}_k \mathbf{z}_{ig})}{\sum_{j=0}^{4} \exp(\boldsymbol{\lambda}_j \mathbf{z}_{ig})}, \tag{6}$$

where ig identifies the subject (see Sec. IV.A), and $\mathbf{z}_{ig}$ is a vector containing the subject's standardized Raven test score, the subject's scores on factors 1–3, and a dummy for whether the subject is own-matched high cognitive ability or own-matched low cognitive ability. The vector $\mathbf{z}_{ig}$ further includes an element that equals unity for all subjects, and therefore, the model includes a full set of learner type-specific intercepts. We impose the identifying normalization $\boldsymbol{\lambda}_1 = \mathbf{0}$. In all other respects, the modified model is identical to the preferred specification in Section IV.A. Using the 270 subjects whose character skills we measured, observed for 10 rounds, we estimate the 30 model parameters: 24 parameters that appear in the type probability functions and six parameters of the t distribution. Section IV.B describes the maximum likelihood estimation routine.

Our discussion focuses on table 11, which shows how cognitive ability and character skills affect the average level k choice rule and the average probability of following either the L0 or L1 choice rule. Looking first at panel A of table 11, we see that both cognitive ability and character skills influence the level k choice rules that our subjects adopt. Linearizing the average marginal effects reported in panel A, a one standard deviation increase in the Raven test score raises the average level k choice rule by 0.26, while a one standard deviation increase in factor 2, which captures an increase in agreeableness and emotional stability, raises the average level k choice rule by 0.16. These are increases of 14 percent and 9 percent, respectively, relative to the baseline average level k choice rule of 1.83. Thus, the impact of cognitive ability on the average level k choice rule is about 65 percent bigger than that of character skills operating through factor 2.

Looking next at panel B of table 11, we see that the Raven test score and factor 2 also significantly decrease (at the 5 percent level) the average probability of following either the L0 or L1 choice rule. Finally, we note that factor 3, which captures openness, extraversion, and CFC, also increases the average level k choice rule (at the 10 percent level) but does not have a statistically significant effect on the average probability of following either the L0 or L1 choice rule.

VI. Conclusion

Cognitive ability and character skills vary greatly across the population. Our analysis shows that cognition and character affect behavior and learning

TABLE 11
EFFECTS OF COGNITIVE ABILITY AND CHARACTER SKILLS ON LEVEL *k* CHOICE RULES

	AVERAGE MARGINAL EFFECTS			
BASELINE	Standardized Raven Test Score (1)	Factor 1 (2)	Factor 2 (3)	Factor 3 (4)
A. Average Level *k* Choice Rule				
1.832***	.264***	−.038	.160**	.131*
(.065)	(.097)	(.067)	(.070)	(.068)
B. Average Probability of Following Either the L0 or L1 Choice Rule				
.422***	−.113**	−.004	−.073**	−.044
(.036)	(.058)	(.038)	(.036)	(.036)

NOTE.—The estimation uses the 270 subjects whose character skills we measured, observed for 10 rounds. We standardized the Raven test scores using these 270 subjects. The baseline average level *k* choice rule was obtained by computing the average level *k* choice rule for each subject, conditional on $\mathbf{z}_{ig}$ (i.e., on the subject's standardized Raven test score, the subject's scores on factors 1–3, and a dummy for whether the subject is own-matched high–cognitive ability or own-matched low–cognitive ability), and then averaging over the subjects. The baseline average probability of following either the L0 or L1 choice rule was obtained by computing the probability that each subject follows either the L0 or L1 choice rule, conditional on $\mathbf{z}_{ig}$, and then averaging over the subjects. The reported marginal effects are averages of the individual-level marginal effects (obtained using differentiation). The matching group dummy did not have a statistically significant effect on the average level *k* choice rule or on the average probability of following either the L0 or L1 choice rule. Tests for the significance of the baseline quantities are one-sided, while all other tests are two-sided. Standard errors are shown in parentheses.
  * Significant at the 10 percent level.
  ** Significant at the 5 percent level.
  *** Significant at the 1 percent level.

in a strategic environment with repeated interaction: cognitive ability and character skills help to drive observed heterogeneity in choices and earnings and strongly predict how quickly groups of agents learn to play equilibrium.

Our finding that more cognitively able, more agreeable, and more emotionally stable people perform better and learn faster is important for understanding how boundedly rational people might operate in real-world environments that involve some element of strategic interaction. Our results are also relevant when interpreting close-to-equilibrium behavior: even if average behavior mimics equilibrium play quite closely after some period of learning, low–cognitive ability agents might nonetheless be earning substantially less than their high–cognitive ability counterparts, with potentially significant implications for fairness and efficiency.

Our findings raise potentially far-reaching practical and ethical questions. For instance, How much protection should public policy afford to slow learners when they operate in markets, especially newer markets in which some participants have price-setting power? How can the design of institutions and mechanisms take into account the impact of bounded

rationality on how agents learn to behave in the strategic environment implied by the rules of the institution or mechanism? Is redistribution appropriate to correct for differences in outcomes when people of different cognitive abilities and character skills interact repeatedly? Our findings also link to a wider debate about how clever agents could create environments and mechanisms designed to exploit learning deficiencies (Sobel 2000, 259). In our context, more cognitively able agents may expend resources to ensure that they interact with those less cognitively able. In addition to being socially wasteful, such efforts have the potential to increase income inequality.

We leave it to future research to investigate the merits of interventions, such as training and advice, legal protections, and redistributive policies, designed to reduce the performance gap between heterogeneous agents who interact repeatedly or to mitigate the impact of differences in learning speed.

Appendix

A. *Effect of Allocation to Cognitive Ability Type*

This paper shows that subjects higher in cognitive ability choose lower numbers and earn more. In this appendix, we consider whether these results could be driven by an effect of allocation to cognitive ability type per se on behavior.

First, we note that the within–matching group effects of cognitive ability on behavior that we describe in Section V cannot be driven by differences in the allocation to cognitive ability type. In particular, within the set of subjects who were matched into groups of three high–cognitive ability subjects, an increase in cognitive ability influences behavior by reducing mean choices (the effect is substantial and significant at the 1 percent level; see Sec. V.B.2). These own-matched high-ability subjects all share the same cognitive ability type allocation.

Next, we show that the allocation to cognitive ability type did not have a statistically significant effect on behavior. Thus, we find no evidence that the allocation to cognitive ability type encouraged or discouraged effort. Recall that each subject was allocated to cognitive ability type according to whether her test score was in the top or bottom half of the distribution of scores in her session (see Sec. II.B). To test whether the allocation to cognitive ability type per se affected behavior, we regress mean p beauty contest choices and earnings across the 10 rounds on cognitive ability type controlling for a subject's own test score by including a full set of Raven test score dummies (this allows a fully flexible nonlinear control for a subject's own cognitive ability). We use only the 445 subjects whose Raven test scores lie in the range of overlap where, conditioning on a given score, some subjects were allocated to the high–cognitive ability type and others to the low-ability type. The range of overlap arises because the session median Raven test score varied from 37.0 to 43.5.[35]

---

[35] Figure 1*B* in Sec. II.B shows the densities of Raven test scores for high- and low-ability subjects separately and thus illustrates the range of overlap in the middle cognitive ability range.

Table A1 reports the results of these regressions. The regressions in the "any" column use all the 445 subjects whose Raven test scores lie in the range of overlap. Since the matching procedure implies that the allocation to cognitive ability type influenced the proportion of high-ability opponents on average, these regressions control for the cognitive ability of opponents by including a full set of dummies for the proportion (1, 0.5, or 0) of high-ability opponents. The next three columns repeat the exercise, restricting attention to subjects facing a given proportion of high-ability opponents (the sum of such subjects is less than 445 because the three ranges of overlap calculated conditional on the proportion of high-ability opponents are different from each other, with the union given by the unconditional range of overlap).

Looking at column 1 in table A1, we see that being allocated to the high-ability type increases choices and earnings on average, but the effects are not statistically significant. The next three columns, which condition on the proportion of high-ability opponents, also show no statistically significant effects. We also note that the signs of the estimated coefficients vary and that the magnitudes are not monotonic in the proportion of high-ability opponents. Table A2 shows similar results when we restrict attention only to the first round or to the second round.

Recall that the mean choice of high–cognitive ability subjects is 2.2 lower than that of low-ability subjects (see Sec. III.B.1). Monte Carlo simulations show that if half of this difference was due to a negative effect on choices of being allocated

TABLE A1
Effect of Allocation to High Cognitive Ability Type (All Rounds)

	Proportion of High–Cognitive Ability Opponents			
	Any (1)	1 (2)	.5 (3)	0 (4)
Effect on mean choices	.978	.737	2.305	−.490
	(.957)	(1.383)	(1.405)	(2.022)
Effect on earnings ($ per round)	.143	−.020	−.071	.266
	(.139)	(.237)	(.257)	(.211)
Subject-round observations	4,450	1,070	610	900

Note.—The models are linear and are estimated using OLS regressions. All models include a full set of Raven test score dummies, which control for a subject's own cognitive ability, and an intercept. The regressions in col. 1 use the 445 subjects (observed for 10 rounds) whose Raven test scores lie in the range of overlap where, conditioning on a given score, some subjects were allocated to the high–cognitive ability type and others to the low-ability type (see Sec. II.B). These regressions control for the cognitive ability of opponents using a full set of dummies for the proportion (1, 0.5, or 0) of high-ability opponents. Columns 2–4 repeat the exercise, restricting attention to subjects facing a given proportion of high-ability opponents (the sum of such subjects is less than 445 because the three ranges of overlap calculated conditional on the proportion of high-ability opponents are different from each other, with the union given by the unconditional range of overlap). The choice regressions control for previous group behavior by including the mean group choice in the previous round and the square of the mean group choice in the previous round. Heteroskedasticity-consistent standard errors, with clustering at the group level, are shown in parentheses.
  * Significant at the 10 percent level (two-sided tests).
  ** Significant at the 5 percent level (two-sided tests).
  *** Significant at the 1 percent level (two-sided tests).

TABLE A2
EFFECT OF ALLOCATION TO HIGH–COGNITIVE ABILITY TYPE (Rounds 1 and 2)

	PROPORTION OF HIGH–COGNITIVE ABILITY OPPONENTS			
	Any (1)	1 (2)	.5 (3)	0 (4)
	A. Round 1			
Effect on mean choices	3.815	.607	4.527	7.268*
	(2.672)	(3.976)	(6.461)	(4.103)
Effect on earnings ($)	.078	−.062	.422	−.201
	(.406)	(.612)	(.809)	(.629)
Subject-round observations	445	107	61	90
	B. Round 2			
Effect on mean choices	.606	5.517	.477	−5.025
	(2.331)	(3.949)	(5.496)	(3.095)
Effect on earnings ($)	.041	−.701	.274	.543
	(.421)	(.673)	(.717)	(.702)
Subject-round observations	445	107	61	90

NOTE.—See the notes to table A1. Subjects are observed for a single round (round 1 or round 2). Heteroskedasticity-consistent standard errors, with clustering at the group level (apart from the round 1 choice regressions, since within-group first-round choices are independent), are shown in parentheses.
* Significant at the 10 percent level (two-sided tests).
** Significant at the 5 percent level (two-sided tests).
*** Significant at the 1 percent level (two-sided tests).

to the high-ability type, then there would be a less than 1 percent chance of estimating an effect of being allocated to the high-ability type that is greater than or equal to our actual estimate of 0.978.[36]

Finally, we note that the range of overlap is sufficiently large to detect effects of cognitive ability. For instance, if we estimate the effect of cognitive ability on choices within the 108 own-matched high-ability subjects whose Raven test scores lie in the range of overlap, we find a substantial negative effect of cognitive ability on choices, significant at the 1 percent level, thus replicating the finding from Section V.B.2 described in the second paragraph of this section.

B. Further Goodness of Fit Analysis

Tables A3 and A4 provide further evidence that our structural level k model of learning fits well. Table A3 reports values of log likelihoods and of sums of squared deviations of choices, earnings, and convergence statistics computed from simulated choices, which together show that our model fits the observed data well. The table also provides evidence from Vuong tests (for nonnested model comparisons) or likelihood ratio tests (for nested model comparisons) that our preferred specification fits better than various alternative specifications. We now provide a brief description of each of the alternative specifications:

[36] The Monte Carlo simulations use a data generation process based on the observed data.

	Preferred Specification	ALTERNATIVE SPECIFICATIONS			
		AS1	AS2	AS3	AS4
Sum of squared deviations of choices:					
Own-matched high-ability subjects	17.1	55.5	16.8	32.0	14.6
Own-matched low-ability subjects	30.6	125.0	39.2	57.0	35.1
Cross-matched high-ability subjects	22.5	84.4	19.7	48.2	16.1
Cross-matched low-ability subjects	24.7	53.7	51.3	29.0	13.4
Sum of squared deviations of earnings:					
Cross-matched high-ability subjects	.190	.220	.269	.224	.287
Sum of squared deviations of convergence statistics:					
Own-matched high-ability groups	2.26	.47	1.57	2.82	.95
Cross-matched groups	.53	.08	1.77	.71	.43
Own-matched low-ability groups	.12	.81	8.32	.24	.17
Log likelihood	−21,510	−21,563	−23,302	−21,780	−21,556
Test against the preferred specification: p-value	. . .	.007[a]	.000[b]	.000[a]	.054[a]

	ALTERNATIVE SPECIFICATIONS			
	AS5	AS6	AS7	AS8
Sum of squared deviations of choices:				
Own-matched high-ability subjects	40.8	28.5	12.2	755.2
Own-matched low-ability subjects	76.6	75.5	32.7	435.3
Cross-matched high-ability subjects	44.5	15.7	34.9	574.4
Cross-matched low-ability subjects	42.5	30.1	19.3	406.8
Sum of squared deviations of earnings:				
Cross-matched high-ability subjects	.246	.481	.216	.585
Sum of squared deviations of convergence statistics:				
Own-matched high-ability groups	3.43	9.90	4.16	57.73
Cross-matched groups	.88	.10	.07	79.40
Own-matched low-ability groups	.32	.46	.38	17.43
Log likelihood	−21,599	−21,534	−21,522	−26,142
Test against the preferred specification: p-value	.000[b]	.003[b]	.104[b]	.000[a]

NOTE.—Descriptions of each alternative specification are in Sec. B of the Appendix. Squared deviations of choices and earnings (in dollars) were computed in each round from the simulated choices generated as described in the notes to fig. 7 and then summed over rounds. By construction, the sum of squared deviations of earnings (i) is identical for cross-matched high-ability and low-ability subjects and (ii) is zero for both own-matched high-ability and low-ability subjects. Squared deviations of convergence statistics were computed for each of the three proportions of equilibrium and close-to-equilibrium play in table 3 and then summed and multiplied by 100.

[a] Two-sided Vuong test.
[b] Likelihood ratio test.

TABLE A4
GOODNESS OF FIT: IN-SAMPLE VERSUS OUT-OF-SAMPLE

	GOODNESS OF FIT: ESTIMATION USING ALL ROUNDS		GOODNESS OF FIT: ESTIMATION USING ROUNDS 1–8	
	All Rounds (1)	Rounds 9 and 10 (2)	All Rounds (3)	Rounds 9 and 10 (4)
Sum of squared deviations of choices:				
Own-matched high-ability subjects	17.1	2.48	12.7	.36
Own-matched low-ability subjects	30.6	1.17	34.3	4.07
Cross-matched high-ability subjects	22.5	1.17	25.1	2.24
Cross-matched low-ability subjects	24.7	1.11	18.5	2.08
Sum of squared deviations of earnings:				
Cross-matched high-ability subjects	.190	.049	.345	.154
Sum of squared deviations of convergence statistics:				
Own-matched high-ability groups	. . .	2.26	. . .	.50
Cross-matched groups	. . .	.53	. . .	.05
Own-matched low-ability groups	. . .	.12	. . .	.13
Log likelihood	−21,510	. . .	−21,553	. . .

NOTE.—The statistics were computed in the same way as described in table A3, except that cols. 2 and 4 report the statistics computed only for rounds 9 and 10, while cols. 3 and 4 were computed from simulated choices based on an estimate of the parameter vector θ obtained using only the data from rounds 1–8. The statistics in col. 4 are therefore out-of-sample quantities.

- AS1: Once an Lk–$(k + 1)$ rule learner switches to the level $(k + 1)$ choice rule, she never switches back. In each round $r \geq 3$, one-eighth of the Lk–$(k + 1)$ rules learners switch to the level $(k + 1)$ choice rule; therefore, as in the preferred specification, in round 2 they all follow the level k choice rule and by round 10 they all follow the Lk–$(k + 1)$ choice rule.
- AS2: The normal distribution replaces the t distribution in the model of the choice process.
- AS3: The choices of 100 come from the same t distribution as for the other choices (instead of the probability of a choice of 100 coming from the round-specific but cognitive ability independent empirical frequency observed in the sample).
- AS4: Subjects following the level k choice rule for $k > 0$ take into account the effect of their own choice on the target and understand that subjects following lower level k choice rules do so as well. For $k > 0$, $\mu(k, \mathbf{x}_{g,r-1})$ is thus given by a lower fraction of $\overline{x}_{g,r-1}$ than in the preferred specification.
- AS5: Rule learner types are not included in the model (so there are just five standard level k learner types, with $k \in \{0, 1, 2, 3, 4\}$, who follow the level k choice rule in all rounds $r \geq 2$).
- AS6: The probabilities of learner types are not allowed to vary with subjects' own cognitive ability or with whether subjects are in own-matched or cross-matched groups.
- AS7: The probabilities of learner types are not allowed to vary with whether subjects are in own-matched or cross-matched groups (but are allowed to vary with subjects' own cognitive ability).

- AS8: The experience-weighted attraction (EWA) learning model of Camerer and Ho (1999) replaces the level k model of learning.[37]

Table A4 shows that our structural level k model of learning continues to perform well out of sample. In particular, the table reports how the model performs when we simulate choices for all rounds but estimate the model's parameters using only the data from rounds 1–8 (cols. 3 and 4), compared to performance using parameters estimated from all the data (cols. 1 and 2). In each case, we provide statistics of fit computed from the simulated choices for all rounds and for rounds 9 and 10 alone.

C. Disaggregating Choices and Earnings Statistics for Cross-Matched Subjects

In Section III.B.3 we found that within cross-matched groups, high- and low-ability subjects choose similarly on average and high-ability subjects earn more than low-ability subjects. Recall that a cross-matched group is made up of either two high-ability subjects and one low-ability subject or one high-ability subject and two low-ability subjects. Figure A1A shows that, in both types of cross-matched group, high- and low-ability subjects choose similarly on average. More formally, over the 10 rounds, the difference between the average choices of high- and low-ability subjects in cross-matched groups with two high-ability subjects is not statistically significant (two-sided $p = .543$).[38] Similarly, over the 10 rounds, the difference between the average choices of high- and low-ability subjects in cross-matched groups with one high-ability subject is not statistically significant (two-sided $p = .286$).

---

[37] We "burn in" the initial attractions using the choice probabilities given by an ordered probit model estimated on the first-round data (for discussion see Ho, Camerer, and Chong [2007, fn. 6]). We set the initial experience weight, $N(0)$, to unity: Ho et al. (182) impose the same restriction, while noting that the influence of this parameter decreases rapidly over rounds and most experimental subjects have weak priors. We estimate a total of 16 parameters: i.e., the four remaining parameters, ϕ, λ, ρ, and δ, for own-matched high–cognitive ability subjects, cross-matched high-ability subjects, cross-matched low-ability subjects, and own-matched low-ability subjects. The poor fit of the EWA model shown in table A3 (AS8) is consistent with previous estimations of the EWA model using beauty contest data (see Camerer and Ho [1999, table 1] and the discussion at pp. 863–64). Despite the poor fit of the EWA model, we briefly discuss differences in parameter estimates by cognitive ability type. High-ability subjects have a statistically significantly lower discount factor for previous attractions, ϕ, and a statistically significantly higher attraction sensitivity, λ. We find no statistically significant differences by cognitive ability type in the discount factor for experience, ρ, or in the relative weight given to belief-based learning vs. reinforcement learning in updating attractions, δ. The latter result is perhaps not surprising because whenever beliefs are given nonzero weight, i.e., $\delta > 0$, attraction updating requires a subject to know her own payoff in the previous round and to compute all forgone payoffs. In other words, provided that $\delta > 0$, the information required to follow the EWA learning rule does not depend on the value of δ.

[38] Throughout this section of the Appendix, all statistical tests are based on OLS regressions and use heteroskedasticity-consistent standard errors. We cluster at the group level to allow for within-group nonindependence. The regressions use 10 rounds of observations on either the 165 subjects in cross-matched groups with two high-ability subjects or the 165 subjects in cross-matched groups with one high-ability subject.

Figure A1*B* shows that, in both types of cross-matched groups, high-ability subjects tend to earn more than low-ability subjects. Over the 10 rounds, high-ability subjects in cross-matched groups with two high-ability subjects earn $2.92 more than low-ability subjects in the same type of cross-matched group (two-sided $p = .173$). Similarly, high-ability subjects in cross-matched groups with one high-ability subject earn $5.10 more than low-ability subjects in the same type of cross-matched group (two-sided $p = .013$).

D. Disaggregating Convergence Statistics for Cross-Matched Groups

Table A5 shows the frequencies of equilibrium and close-to-equilibrium play in the final two rounds of the experiment separately for cross-matched groups
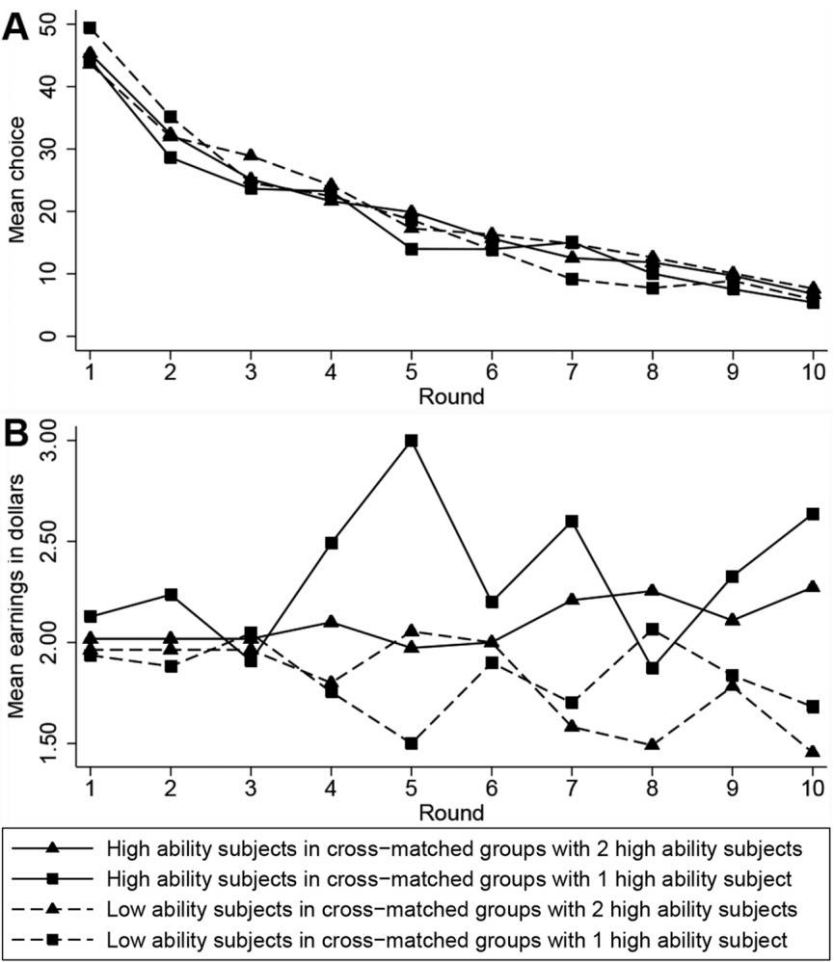

FIG. A1.—Disaggregated *p* beauty contest choices and earnings of cross-matched subjects. *A*, Round-by-round means of choices. *B*, Round-by-round means of earnings.

TABLE A5

PROPORTIONS OF EQUILIBRIUM AND CLOSE-TO-EQUILIBRIUM PLAY IN ROUNDS 9 AND 10,
SPLIT BY TYPE OF CROSS-MATCHED GROUP

	Equilibrium (1)	Group Mean ≤1 (2)	Group Mean ≤2 (3)
Observed proportions:			
Cross-matched groups with two high-ability subjects	.164	.282	.345
Cross-matched groups with one high-ability subject	.127	.282	.391
Cross-group differences in proportions:			
Cross-matched groups with two high-ability subjects vs. cross-matched groups with one high-ability subject	.036	.000	−.045
	(.065)	(.081)	(.086)

NOTE.—The table is based on 220 group-round observations (55 cross-matched groups with two high-ability subjects and 55 cross-matched groups with one high-ability subject, observed in rounds 9 and 10). Column 1 reports the proportion of group-round observations in rounds 9 and 10 in which all three group members choose 0. Column 2 (3) reports the proportion of group-round observations in rounds 9 and 10 in which the mean choice of the three group members is less than or equal to 1 (2). The statistical tests of differences in proportions are based on OLS regressions. Heteroskedasticity-consistent standard errors (with clustering at the group level) are shown in parentheses.

  * Significant at the 10 percent level (two-sided tests).
  ** Significant at the 5 percent level (two-sided tests).
  *** Significant at the 1 percent level (two-sided tests).

with two high-ability subjects and for cross-matched groups with one high-ability subject. We note that there are no statistically significant differences between the two types of cross-matched groups in the frequencies of equilibrium or close-to-equilibrium play.

E. Further Evidence on the Interplay of Cognitive Ability and Character Skills

In Section V.B.2, we showed that (i) both cognitive ability and character skills influence behavior and earnings in our repeated strategic interaction, and (ii) the estimated influence of cognitive ability does not depend on whether we control for character skills and vice versa. Here, we dig deeper into the interplay of cognitive ability and character skills. First, we extend the analysis of Section V.B.2 to allow interactions between cognitive ability and character skills. Second, we investigate whether (*a*) the estimated effect of cognitive ability is robust to introducing character skill dummies and (*b*) the estimated effects of character skills are robust to introducing cognitive ability dummies.

  Panel A of table A6 presents this analysis for behavior. As a benchmark, model 1 in panel A is identical to model 3 in panel A of table 8 in Section V.B.2.

  Model 2 in panel A of table A6 adds interactions between cognitive ability and each of the factors that capture character skills. We see that the estimates of the noninteracted effects of cognitive ability and character skills on average choices across the 10 rounds are close to the estimates in model 1 without interactions. Furthermore, we find no statistically significant interactions between cognitive ability and either factor 2 or factor 3. We do find a weakly statistically

FURTHER RESULTS ON THE EFFECTS OF COGNITIVE ABILITY AND CHARACTER SKILLS
ON MEAN p BEAUTY CONTEST CHOICES AND PER-ROUND EARNINGS

	(1)	(2)	(3)	(4)
	A. Mean p Beauty Contest Choices			
Standardized Raven test score	−2.177**	−2.315**	−2.189**	. . .
	(.985)	(.977)	(.904)	
Factor 1	.153	.187	. . .	.243
	(.530)	(.515)		(.545)
Factor 2	−1.238**	−1.286***	. . .	−1.211**
	(.484)	(.484)		(.499)
Factor 3	.411	.369	. . .	.324
	(.585)	(.574)		(.634)
Standardized Raven test score × factor 1	. . .	−.785*	. . .	. . .
		(.444)		
Standardized Raven test score × factor 2	. . .	.533	. . .	. . .
		(.529)		
Standardized Raven test score × factor 3	. . .	−.281	. . .	. . .
		(.530)		
Raven test score dummies	No	No	No	Yes
Factor 1 dummies	No	No	Yes	No
Factor 2 dummies	No	No	Yes	No
Factor 3 dummies	No	No	Yes	No
Subject-round observations	2,700	2,700	2,700	2,700
	B. Earnings ($ per Round)			
Standardized Raven test score	.113*	.124*	.135*	. . .
	(.067)	(.064)	(.070)	
Factor 1	.031	.031	. . .	.053
	(.051)	(.051)		(.050)
Factor 2	.129**	.134**	. . .	.156***
	(.051)	(.051)		(.059)
Factor 3	−.048	−.049	. . .	−.065
	(.047)	(.047)		(.048)
Standardized Raven test score × factor 1	. . .	.045	. . .	. . .
		(.036)		
Standardized Raven test score × factor 2	. . .	−.008	. . .	. . .
		(.048)		
Standardized Raven test score × factor 3	. . .	−.030	. . .	. . .
		(.046)		
Raven test score dummies	No	No	No	Yes
Factor 1 dummies	No	No	Yes	No
Factor 2 dummies	No	No	Yes	No
Factor 3 dummies	No	No	Yes	No
Subject-round observations	2,700	2,700	2,700	2,700

NOTE.—The models are linear and are estimated using OLS regressions. The regressions use the 270 subjects whose character skills we measured, observed for 10 rounds. We standardize the Raven test scores using these 270 subjects. All models include an intercept and a dummy for whether subjects are own-matched high–cognitive ability or own-matched low–cognitive ability. Raven test score dummies are dummies for each unique value of the Raven test score. Factor j dummies for $j = 1, 2, 3$ are dummies for each decile of factor j, i.e., 10 dummies for each factor. Heteroskedasticity-consistent standard errors with clustering at the group level are shown in parentheses.
  * Significant at the 10 percent level (two-sided tests).
  ** Significant at the 5 percent level (two-sided tests).
  *** Significant at the 1 percent level (two-sided tests).

significant negative interaction (two-sided $p = .080$) between cognitive ability and factor 1, which captures conscientiousness, grit, and CFC. That is, we find some evidence that marginal changes in cognitive ability reduce mean choices more for subjects with a higher factor 1 score (who are more conscientious and gritty).

Instead of linear effects of the factors, model 3 in panel A of table A6 includes 30 character skill dummies (10 for each factor). We see that the estimated influence of cognitive ability on behavior is robust to the inclusion of these character skill dummies. Instead of a linear effect of cognitive ability, model 4 in panel A includes a dummy for each of the 61 possible scores in the Raven test. We see that the estimated influence of character skills on behavior is robust to the inclusion of these cognitive ability dummies.

Finally, panel B of table A6 replicates this analysis for earnings instead of behavior. As a benchmark, model 1 in panel B is identical to the model in column 1 of table 9 in Section V.B.2. The results for earnings are similar to those for behavior described above, although we no longer find a statistically significant interaction between cognitive ability and factor 1.

References

Agranov, M., A. Caplin, and C. Tergiman. 2011. "The Process of Choice in Guessing Games." Manuscript, California Inst. Tech.

Agranov, M., E. Potamites, A. Schotter, and C. Tergiman. 2012. "Beliefs and Endogenous Cognitive Levels: An Experimental Study." *Games and Econ. Behavior* 75 (2): 449–63.

Almlund, M., A. L. Duckworth, J. J. Heckman, and T. Kautz. 2011. "Personality Psychology and Economics." In *Handbook of the Economics of Education*, vol. 4, edited by E. Hanushek, S. Machin, and L. Woessman, 1–181. Amsterdam: Elsevier.

Al-Ubaydli, O., G. Jones, and J. Weel. 2013. "Patience, Cognitive Skill, and Coordination in the Repeated Stag Hunt." *J. Neuroscience, Psychology, and Econ.* 6 (2): 71–96.

Arbak, E., and M. C. Villeval. 2013. "Voluntary Leadership: Motivation and Influence." *Soc. Choice and Welfare* 40 (3): 635–62.

Bayer, R. C., and L. Renou. 2012. "Logical Abilities and Behavior in Strategic-Form Games." Manuscript, Univ. Leicester.

Berndt, E., B. Hall, R. Hall, and J. Hausman. 1974. "Estimation and Inference in Nonlinear Statistical Models." *Ann. Econ. and Soc. Measurement* 3 (4): 653–65.

Boone, C., B. De Brabander, M. Carree, G. de Jong, W. van Olffen, and A. van Witteloostuijn. 2002. "Locus of Control and Learning to Cooperate in a Prisoner's Dilemma Game." *Personality and Individual Differences* 32 (5): 929–46.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. ter Weel. 2008. "The Economics and Psychology of Personality Traits." *J. Human Resources* 43 (4): 972–1059.

Borghans, L., B. H. H. Golsteyn, J. J. Heckman, and J. E. Humphries. 2011. "Identification Problems in Personality Psychology." *Personality and Individual Differences* 51 (3): 315–20.

Brañas-Garza, P., T. García-Muño, and R. Hernán. 2012. "Cognitive Effort in the Beauty Contest Game." *J. Econ. Behavior and Org.* 83 (2): 254–60.

Brandstätter, H., and M. Königstein. 2001. "Personality Influences on Ultimatum Bargaining Decisions." *European J. Personality* 15 (S1): S53–S70.

Brocas, I., J. D. Carrillo, S. W. Wang, and C. F. Camerer. 2014. "Imperfect Choice or Imperfect Attention? Understanding Strategic Thinking in Private Information Games." *Rev. Econ. Studies* 81 (3): 944–70.

Burks, S. V., J. P. Carpenter, L. Goette, and A. Rustichini. 2009. "Cognitive Skills Affect Economic Preferences, Strategic Behavior, and Job Attachment." *Proc. Nat. Acad. Sci.* 106 (19): 7745–50.

Burks, S. V., C. Lewis, P. Kivi, et al. 2014. "Moving Ahead by Thinking Backwards: Cognitive Skills, Personality, and Economic Preferences in Collegiate Success." IZA Discussion Paper no. 7952, Inst. Study Labor, Bonn.

Burnham, T., D. Cesarini, M. Johannesson, P. Lichtenstein, and B. Wallace. 2009. "Higher Cognitive Ability Is Associated with Lower Entries in a *p*-Beauty Contest." *J. Econ. Behavior and Org.* 72 (1): 171–75

Camerer, C. F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton, NJ: Princeton Univ. Press.

Camerer, C. F., and T. H. Ho. 1999. "Experience-Weighted Attraction Learning in Normal Form Games." *Econometrica* 67 (4): 827–74.

Camerer, C. F., T. H. Ho, and J. K. Chong. 2004. "A Cognitive Hierarchy Model of Games." *Q.J.E.* 119 (3): 861–98.

Cardella, E. 2012. "Learning to Make Better Strategic Decisions." *J. Econ. Behavior and Org.* 84 (1): 382–92.

Carpenter, P. A., M. A. Just, and P. Shell. 1990. "What One Intelligence Test Measures: A Theoretical Account of the Processing in the Raven Progressive Matrices Test." *Psychological Rev.* 97 (3): 404–13.

Charness, G., A. Rustichini, and J. van de Ven. 2011. "Self-Confidence and Strategic Deterrence." Manuscript, Univ. California, Santa Barbara.

Coricelli, G., and R. Nagel. 2009. "Neural Correlates of Depth of Strategic Reasoning in Medial Prefrontal Cortex." *Proc. Nat. Acad. Sci.* 106 (23): 9163–68.

Costa-Gomes, M. A., and V. P. Crawford. 2006. "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study." *A.E.R.* 96 (5): 1737–68.

Costa-Gomes, M. A., V. P. Crawford, and B. Broseta. 2001. "Cognition and Behavior in Normal-Form Games: An Experimental Study." *Econometrica* 69 (5): 1193–1235.

Costa-Gomes, M. A., V. P. Crawford, and N. Iriberri. 2009. "Comparing Models of Strategic Thinking in Van Huyck, Battalio, and Beil's Coordination Games." *J. European Econ. Assoc.* 7 (2–3): 365–76.

Costa-Gomes, M. A., and G. Weizsäcker. 2008. "Stated Beliefs and Play in Normal-Form Games." *Rev. Econ. Studies* 75 (3): 729–62.

Crawford, V. P., M. A. Costa-Gomes, and N. Iriberri. 2013. "Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications." *J. Econ. Literature* 51 (1): 5–62.

Crawford, V. P., and N. Iriberri. 2007a. "Fatal Attraction: Salience, Naïveté, and Sophistication in Experimental 'Hide-and-Seek' Games." *A.E.R.* 97 (5): 1731–50.

———. 2007b. "Level-*k* Auctions: Can a Nonequilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions?" *Econometrica* 75 (6): 1721–70.

Cunha, F., J. J. Heckman, and S. M. Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica* 78 (3): 883–931.

Daly, M., C. P. Harmon, and L. Delaney. 2009. "Psychological and Biological Foundations of Time Preference." *J. European Econ. Assoc.* 7 (2–3): 659–69.

Delaney, L., C. Harmon, and M. Ryan. 2013. "The Role of Noncognitive Traits in Undergraduate Study Behaviours." *Econ. Educ. Rev.* 32:181–95.

DeYoung, C. G., J. B. Hirsh, M. S. Shane, X. Papademetris, N. Rajeevan, and J. R. Gray. 2010. "Testing Predictions from Personality Neuroscience Brain Structure and the Big Five." *Psychological Sci.* 21 (6): 820–28.

Duckworth, A. L., C. Peterson, M. D. Matthews, and D. R. Kelly. 2007. "Grit: Perseverance and Passion for Long-Term Goals." *J. Personality and Soc. Psychology* 92 (6): 1087–1101.

Duffy, J., and R. Nagel. 1997. "On the Robustness of Behaviour in Experimental 'Beauty Contest' Games." *Econ. J.* 107 (445): 1684–1700.

Dufwenberg, M., R. Sundaram, and D. J. Butler. 2010. "Epiphany in the Game of 21." *J. Econ. Behavior and Org.* 75 (2): 132–43.

Eysenck, M. W. 1985. "Anxiety and Cognitive-Task Performance." *Personality and Individual Differences* 6 (5): 579–86.

Fahr, R., and B. Irlenbusch. 2008. "Identifying Personality Traits to Enhance Trust between Organisations: An Experimental Approach." *Managerial and Decision Econ.* 29 (6): 469–87.

Fischbacher, U. 2007. "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Econ.* 10 (2): 171–78.

Fouarge, D., T. Schils, and A. de Grip. 2013. "Why Do Low-Educated Workers Invest Less in Further Training?" *Appl. Econ.* 45 (18): 2587–2601.

Frederick, S. 2005. "Cognitive Reflection and Decision Making." *J. Econ. Perspectives* 19 (4): 25–42.

Gneezy, U., A. Rustichini, and A. Vostroknutov. 2010. "Experience and Insight in the Race Game." *J. Econ. Behavior and Org.* 75 (2): 144–55.

Goldfarb, A., and M. Xiao. 2011. "Who Thinks about the Competition? Managerial Ability and Strategic Entry in US Local Telephone Markets." *A.E.R.* 101 (7): 3130–61.

Gray, J., and P. Thompson. 2004. "Neurobiology of Intelligence: Science and Ethics." *Nature Revs. Neuroscience* 5 (6): 471–82.

Heckman, J. J., J. E. Humphries, and G. Veramendi. 2015. "The Causal Effects of Education on Earnings and Health." Manuscript, Univ. Chicago.

Heckman, J. J., and T. Kautz. 2012. "Hard Evidence on Soft Skills." *Labour Econ.* 19 (4): 451–64.

———. 2014. "Achievement Tests and the Role of Character in American Life." In *The Myth of Achievement Tests: The GED and the Role of Character in American Life*, edited by J. J. Heckman, J. E. Humphries, and T. Kautz. Chicago: Univ. Chicago Press.

Heckman, J. J., J. Stixrud, and S. Urzua. 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *J. Labor Econ.* 24 (3): 411–82.

Ho, T. H., C. F. Camerer, and J. K. Chong. 2007. "Self-Tuning Experience Weighted Attraction Learning in Games." *J. Econ. Theory* 133 (1): 177–98.

Ho, T. H., C. F. Camerer, and K. Weigelt. 1998. "Iterated Dominance and Iterated Best Response in Experimental '*p*-Beauty Contests.'" *A.E.R.* 88 (4): 947–69.

Houser, D., and J. Winter. 2004. "How Do Behavioral Assumptions Affect Structural Inference? Evidence from a Laboratory Experiment." *J. Bus. and Econ. Statis.* 22 (1): 64–79.

John, O. P., E. M. Donahue, and R. L. Kentle. 1991. "The Big Five Inventory: Versions 4a and 54." Manuscript, Inst. Personality and Soc. Res., Univ. California, Berkeley.

John, O. P., L. P. Naumann, and C. J. Soto. 2008. "Paradigm Shift to the Integrative Big Five Trait Taxonomy: History, Measurement, and Conceptual Issues." In *Handbook of Personality: Theory and Research*, edited by O. P. John, R. W. Robins, and L. A. Pervin, 114–58. New York: Guilford.

Johnson, M. K., A. Rustichini, and A. G. MacDonald III. 2009. "Suspicious Personality Predicts Behavior on a Social Decision-Making Task." *Personality and Individual Differences* 47 (1): 30–35.

Jolliffe, I. T. 1995. "Rotation of Principal Components: Choice of Normalization Constraints." *J. Appl. Statis.* 22 (1): 29–35.

Kaiser, H. F. 1960. "The Application of Electronic Computers to Factor Analysis." *Educ. and Psychological Measurement* 20 (1): 141–51.

Kurzban, R., and D. Houser. 2001. "Individual Differences in Cooperation in a Circular Public Goods Game." *European J. Personality* 15 (S1): S37–S52.

Le Coq, C., and J. T. Sturluson. 2012. "Does Opponents' Experience Matter? Experimental Evidence from a Quantity Precommitment Game." *J. Econ. Behavior and Org.* 84 (1): 265–77.

Levitt, S. D., J. A. List, and S. E. Sadoff. 2011. "Checkmate: Exploring Backward Induction among Chess Players." *A.E.R.* 101 (2): 975–90.

Lindqvist, E., and R. Vestman. 2011. "The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment." *American Econ. J.: Appl. Econ.* 2 (1): 101–28.

López, R. 2001. "On *p*-Beauty Contest Integer Games." Econ. and Bus. Working Paper no. 608, Univ. Pompeu Fabra.

Millet, K., and S. Dewitte. 2007. "Altruistic Behavior as a Costly Signal of General Intelligence." *J. Res. Personality* 41 (2): 316–26.

Mueller, G., and E. Plug. 2006. "Estimating the Effect of Personality on Male and Female Earnings." *Indus. and Labor Relations Rev.* 60 (1): 3–22.

Müller, J., and C. Schwieren. 2012. "What Can the Big Five Personality Factors Contribute to Explain Small-Scale Economic Behavior?" Discussion Paper no. 2012-028/1, Tinbergen Inst.

Nagel, R. 1995. "Unraveling in Guessing Games: An Experimental Study." *A.E.R.* 85 (5): 1313–26.

Neeman, Z., T. Kugler, and N. Vulkan. 2013. "Personality Traits and Strategic Behavior: Anxiousness, Aggressiveness and Entry Games." Discussion Paper no. 7-13, Pinhas Sapir Center Development, Tel Aviv Univ.

Nettle, D., and B. Liddle. 2008. "Agreeableness Is Related to Social-Cognitive, but Not Social-Perceptual, Theory of Mind." *European J. Personality* 22 (4): 323–35.

Offerman, T., J. Potters, and J. Sonnemans. 2002. "Imitation and Belief Learning in an Oligopoly Experiment." *Rev. Econ. Studies* 69 (4): 973–97.

Östling, R., J. T. Wang, E. Y. Chou, and C. F. Camerer. 2011. "Testing Game Theory in the Field: Swedish LUPI Lottery Games." *American Econ. J.: Microeconomics* 3 (3): 1–33.

Palacios-Huerta, I., and O. Volij. 2009. "Field Centipedes." *A.E.R.* 99 (4): 1619–35.

Raven, J., J. C. Raven, and J. H. Court. 2000. *Manual for Raven's Progressive Matrices and Vocabulary Scales Section 3: Standard Progressive Matrices.* San Antonio: Pearson.

Rustichini, A., C. G. DeYoung, J. Anderson, and S. V. Burks. 2012. "Toward the Integration of Personality Theory and Decision Theory in the Explanation of Economic and Health Behavior." IZA Discussion Paper no. 6750, Inst. Study Labor, Bonn.

Rydval, O., A. Ortmann, and M. Ostatnicky. 2009. "Three Very Simple Games and What It Takes to Solve Them." *J. Econ. Behavior and Org.* 72 (1): 589–601.

Schnusenberg, O., and A. Gallo. 2011. "On Cognitive Ability and Learning in a Beauty Contest." *J. Econ. Educators* 11 (1): 13–24.

Sobel, J. 2000. "Economists' Models of Learning." *J. Econ. Theory* 94 (2): 241–61.

Stahl, D. O. 1996. "Boundedly Rational Rule Learning in a Guessing Game." *Games and Econ. Behavior* 16 (2): 303–30.

Stahl, D. O., and P. W. Wilson. 1995. "On Players' Models of Other Players: Theory and Experimental Evidence." *Games and Econ. Behavior* 10 (1): 218–54.

Strathman, A., F. Gleicher, D. S. Boninger, and C. Scott Edwards. 1994. "The Consideration of Future Consequences: Weighing Immediate and Distant Outcomes of Behavior." *J. Personality and Soc. Psychology* 66 (4): 742–52.

Szymura, B., and Z. Wodniecka. 2003. "What Really Bothers Neurotics? In Search for Factors Impairing Attentional Performance." *Personality and Individual Differences* 34 (1): 109–26.

Volk, S., C. Thöni, and W. Ruigrok. 2011. "Personality, Personal Values and Cooperation Preferences in Public Goods Games: A Longitudinal Study." *Personality and Individual Differences* 50 (6): 810–15.

Webley, P., and E. K. Nyhus. 2013. "Economic Socialization, Saving and Assets in European Young Adults." *Econ. Educ. Rev.* 33:19–30.