Journal of Economic Behavior & Organization 83 (2012) 254–260

Contents lists available at SciVerse ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

Cognitive effort in the Beauty Contest Game☆

Pablo Brañas-Garza[a,*], Teresa García-Muñoz[a], Roberto Hernán González[a,b]

[a] GLOBE: U. de Granada, Spain
[b] ESI Chapman University, USA

ARTICLE INFO

Article history:
Received 12 July 2011
Received in revised form 30 April 2012
Accepted 25 May 2012
Available online 9 June 2012

Keywords:
Beauty Contest Game
Raven
Cognitive Reflection Test

ABSTRACT

This paper analyzes cognitive effort in 6 different oneshot p-beauty games. We use both Raven and Cognitive Reflection tests to identify subjects' abilities. We find that the Raven test does not provide any insight on Beauty Contest Game playing but CRT does: subjects with higher scores on this test are more prone to play dominant strategies. The results are confirmed when levels of reasoning instead of entries in the BCG are used.

1. Introduction

Recent papers connect individuals' cognitive abilities with performance in different games through different tests (see for instance Burnham et al., 2009; Oechssler et al., 2009; Brañas-Garza et al., 2011). This paper expands on this literature using both the Raven and the Cognitive Reflection Test (CRT hereafter) to study how people play a series of six p-Beauty Contest Games. We find that the Raven test lacks explanatory power, but the CRT makes a difference.

An increasing amount of literature analyzes the connection between economic behavior and cognitive abilities. Frederick (2005) shows that subjects who score high on the CRT are more patient and more willing to take risks in gains. Benjamin et al. (forthcoming) show similar results for Chilean high school students and, with a more heterogeneous sample, Dohmen et al. (2010) also find that cognitive abilities are related to time and risk preferences. Interestingly, Brañas-Garza et al. (2008) find that risk attitudes are similar across subjects with different computational abilities. Oechssler et al. (2009) show that subjects with low scores on a cognitive test are more likely subjected to the conjunction fallacy and to conservatism to update probabilities. Analyzing the entries in a Travelers' dilemma game, Brañas-Garza et al. (2011) find that subjects who score better on a GRE-type math test tend to "undercut" the rival.

Assuming rationality and common knowledge of rationality, the Beauty Contest Game (BCG hereafter) has a unique Nash equilibrium, i.e., play zero. However, this equilibrium has not been observed in the laboratory setting for the one-shot game, although players tend to the equilibrium after several repetitions with feedback. Alternatively, the literature has considered equilibrium strategies according to depths of levels of reasoning (cognitive hierarchy of thinking) that better describe behavior in this game (Nagel, 1995; Camerer et al., 2004). A higher level of reasoning indicates higher strategic behavior by subjects and the belief that rivals are also more strategic.

☆ Financial support from the Spanish Ministry of Science and Innovation (ECO2010-17049, ECO2008-00977/ECON and 2012/00103/001), the Junta of Andalucía (P07.SEJ.02547) and Fundación Ramón Areces (R+D 2011).

* Corresponding author. Tel.: +34 958 246192; fax: +34 958 249995.
E-mail address: pbg@ugr.es (P. Brañas-Garza).

0167-2681/$ – see front matter
http://dx.doi.org/10.1016/j.jebo.2012.05.018

Table 1
Socio-demographic characteristics.

	Mean	St. dev.	Min	Max
(a) Individual				
Age	22.21	2.33	18	31
Female	0.61	0.49	0	1
Health	3.61	0.24	1	5
Risk aversion	4.7	1.39	1	7
Raven	48.91	6.00	12	60
CRT	0.44	0.70	0	3
(b) Household				
Education	2.82	0.23	1	5
Size	4.10	1.03	2	7

Burnham et al. (2009) investigated the relationship between cognitive abilities and choices in a one-shot Beauty Contest Game (BCG) with $p = 1/2$. They found that individuals with higher scores on the cognitive test choose numbers closer to the Nash equilibrium in the one-shot BCG.[1] They point out that this result could be driven by the fact that subjects with lower scores have more mathematical difficulties finding the equilibrium as they choose dominated numbers. But they also argue that this result could be related to differences in predicting other participants' choices (out of the equilibrium). Coricelli and Nagel (2009) show individuals' brain activity is different when playing a BCG with another human participant than when playing with a computer that selects the numbers randomly. Furthermore, they find that subjects with a higher level of reasoning expect other participants to play strategically, while low-level reasoning subjects choose in the belief that others will play randomly (see Coricelli and Nagel, 2009 on the Theory of Mentalizing). According to the Theory of Mentalizing,[2] Bruguier et al. (2010) find that skill in predicting price changes in markets with insider correlates with scores on "Eye Gaze" and "Heider" tests of mentalizing. Interestingly, Bruguier et al. (2010) do not find evidence of correlation between participants ability to predict price changes and their score in a mathematics test.[3]

To sum up, cognitive abilities matters in how people solve games and predict other players' behavior. However, these cognitive abilities do not necessarily reflect higher mathematical capacity. Still, unraveling which cognitive characteristics play a crucial role in rational behavior is an open question.

We analyze the Raven test and the CRT as they have appealing characteristics for playing the BCG. Raven's Progressive Matrices test (Raven et al., 2000) measures visual reasoning and analytic intelligence, the capacity to learn from immediate experience with the problem without relying on previous knowledge, and mathematical reasoning (Mills et al., 1993; Ablard and Mills, 1996). We think that Raven's test is specially relevant in our case as participants in our experiment repeat the BCG six times. Although not feedback is given and the values of the relevant parameter do not remain constant across the games, it may be the case that subjects get some feedback learning (Weber, 2003).

The second test is the CRT proposed by Frederick (2005); a short test with only three brief questions that can be answered in less than 3 minutes. The three items of the CRT are designed such that the intuitive response is incorrect, but can be correctly reconsidered through some deliberation. In this sense, the CRT measures cognitive reflectiveness or impulsiveness, respondents' automatic response versus more elaborate and deliberative thought, and is also correlated with mathematical skills. Anticipation to other players' actions (or types) is a requirement for best response. CRT can help us to distinguish those subjects who are able to elaborate a more sophisticated and less impulsive thinking and, therefore, play accordingly to the Nash equilibrium.

2. Experimental methods

A total of 191 subjects (74 males and 117 females) participated in the experiment. The experiment was run over 8 sessions; 7 sessions with 24 participants each and one session with 23 participants. The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007) at the "old" experimental laboratory of the University of Granada, Spain. The subjects came to the lab and played six rounds of the BCG, one round of the Raven test and one round of the CRT in that order. Subjects were not allowed to use pencils or paper to make calculations. Additionally, they completed some questionnaires and performed some risk lotteries (not reported here).[4]

Table 1 shows the basic descriptive statistics of the sample. *Health* is self-reported health (from bad = 1 to very good = 5), *risk aversion* is the Holt–Laury test (value from 1 to 9).[5]

[1] In a recent paper, Gill and Prowse (2012), the Raven test is used to classify subjects as low/high skilling. In a second step, subjects play BCG within groups created according to these categories (under full information of characteristics).

[2] ". . .*humans detect malevolence or benevolence by online tracking of changes in their environment (rather than, say, logical deduction about the situation at hand)*" (Bruguier et al., 2010, p. 1705).

[3] Coricelli and Nagel (2009) find a similar result.

[4] Subjects responded to the questionnaires and risk lotteries once the Raven test had been completed and before they started the CRT.

[5] The measure of risk aversion was introduced in the models shown along the paper but it was never significant.

Table 2
*M*s by screens.

Screen 1: $M = 2/3$	Screen 4: $M = 1/3$
Screen 2: $M = 1/8$	Screen 5: $M = 1/5$
Screen 3: $M = 3/4$	Screen 6: $M = 1/2$

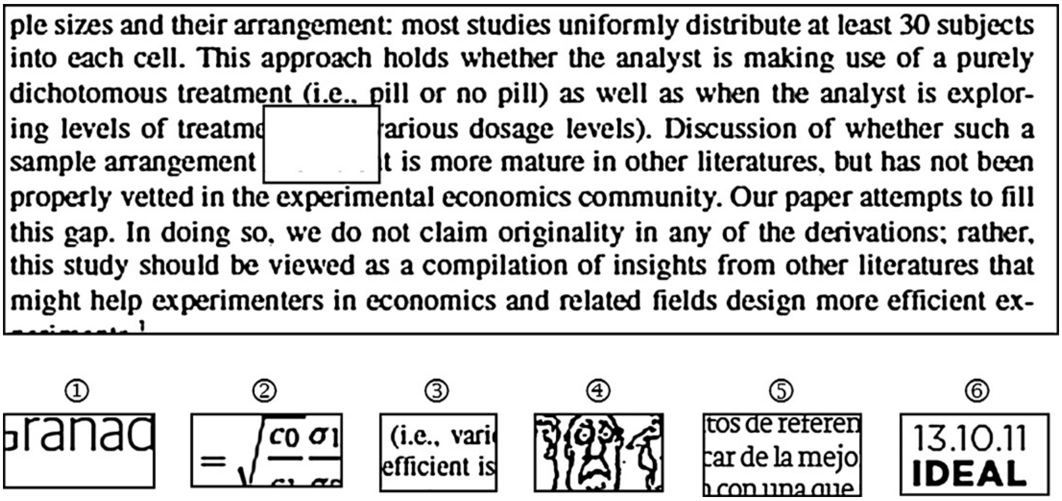

Fig. 1. A figurative example[8] of one item in Raven's test.

The last two variables refers to the household: *size* (number of people living at home) and *education* (max. level of education at home).

2.1. Beauty Contest Game

The Beauty Contest Game[6] consisted of guessing an integer number between 0 and 100 (both limits included) in which the winner is the person whose number is closest to M^* (average of all chosen numbers). In contrast to Burnham et al. (2009), we ran six different one-shot BCG where M – the known multiplier parameter – takes 6 values: 1/8, 1/5, 1/3, 1/2, 2/3 and 3/4.

Playing 0 is the dominant strategy and all players choosing zero is the unique Nash equilibrium. Observe that equilibrium playing does not depend on the value of M. However, the value of M may help to achieve the equilibrium in a lower number of eliminations of dominated strategies.

The subjects were distributed into groups of 24 individuals. The winner of each round received 20 euros. In the event of a tie, the 20 euros were split between those who tied. We did not provide any feedback between trials. Information about the results of the game was provided at the payment stage (see below). As shown in Table 2, all the subjects played the different versions of the game in the same order:

Observe that we chose this particular ordering of values of M in such a way that:

(i) Participants would find it more difficult to learn as the values increase and decrease from one game to the next.
(ii) Furthermore, this design with several repetitions allows us to distinguish between players who play rationally by chance[7] and those thinking about their best strategy.

2.2. Raven test and CRT

Originally developed by Dr. John C. Raven in 1936, Raven's Progressive Matrices are multiple choice tests of abstract reasoning. In each test item, a subject is asked to identify the missing item required to complete a larger pattern (see Fig. 1). In our case, subjects face 60 matrices, that is, they make 60 choices. We calculate $Raven_i$ as the sum of correct answers, hence $Raven_i \in [0, 60]$ where 60 indicates that the subject correctly filled the 60 matrices.

[6] The original instructions are in Spanish. The instructions were provided by Rosemarie Nagel. A copy of the English version is provided in Appendix B.

[7] A subject playing randomly can pick a number close to the Nash equilibrium. However, in contrast with a one-shot game design, the probability of observing repeatedly low entries by this type of subjects is much lower. The lack of monotonic order (increasing/decreasing) makes the problem harder.

[8] We do not reproduce the original figure, but a figurative representation. The essence of the test is finding the missing segment required to complete a larger pattern.

The final score is a measure of ability for abstract analytic reasoning and fluid intelligence, that is, an ability that does not rely on knowledge or skill acquired from experience as opposed to crystallized intelligence (see Horn and Cattell, 1966). Following Burnham et al. (2009), we expected to find a negative[9] relation between high scores on the test and entries in the BCG.

Once the subjects finished the Raven test, they completed the CRT developed by Frederick (2005). The CRT consists of three short questions:

1. A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?
2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?[10]
3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

The three questions have an obvious incorrect answer (10, 100, and 24), that can be easily corrected upon minimal reflection. Those who arrive at the correct answers are less impulsive and more likely to engage in reflective thinking.[11] In this sense, the CRT can be viewed as a combination of cognitive capacity and the disposition for judgement and decision making (Finucane and Gullion, 2010; Toplak et al., 2011). Toplak et al. (2011) put forward that the CRT captures important characteristics of rational thinking that are not measured on other intelligence tests. They argue that humans tend to use the simplest cognitive mechanism, which could mean that sometime behave not fully rational. The CRT is computed as the number of questions answered correctly. Frederick (2005) shows that the scores are highly correlated with some other tests of analytic thinking (such as the ACT, NFC, SAT and WPT). We predicted that subjects that do better on this test are more likely to choose lower entries in the BCG.

Its important to mention here that the subjects completed the CRT as the last task. Moreover, they did the test in front of a computer and without pencils or paper to make their calculations. This may explain why the scores are, on average, not so high as (128 individuals, 67% of the sample pool, did not provide any right answers) compared to those shown in Frederick (2005).[12]

Since we were interested in detecting specific subjects who are able to solve these questions – without any help – this particular set-up posed no problems for us. In this sense, our sample of subjects is a lower bound. Moreover, our BCG is also computerized, hence this appears to be the most “sensible” comparison. As expected, scoring on the Raven test and the CRT are correlated (Spearman’s $\rho = 0.29$; p-value < 0.000).[13]

In addition, males show higher scores in both the Raven test (Cramer’s $\rho = 0.25$; p-value $= 0.002$) and the CRT (Cramer’s $\rho = 0.25$; p-value $= 0.003$).[14]

3. Results

First, we explore the effect of individual cognitive abilities through the BCG. As in previous studies, just a few subjects played according to the Nash equilibrium. For all six games, the choices range from 0 to 100.[15] Fig. 2 shows the mean values of the subjects’ choices, which were classified for their score on the CRT (CRT = 0 versus CRT > 0). It is easy to see that subjects’ choices are related to their performance on the CRT.

Table 3 shows a series of six Tobit models where the dependent variable is, in each case, the individual guess, $g_i \in [0, 100]$. As independent variables we used *female*, $Raven_i$ and CRT_i. The models are presented in the same order in which they were played. There are two salient results: (i) Raven is never significant, and (ii) CRT appears to be significant after two trials. After minimal experience, subjects with a positive score on the CRT behave better.[16]

We study also how subjects play across games. We perform two complementary analysis: (a) panel data pooling observations across the six games; (b) we study the number of times people play according to dominant strategies.

In Table 4 we report the results of a panel data tobit regression (random effects) of the entries. In model A we repeat the analysis performed in Table 3, using panel. In model B we introduce dummies to capture the effect of the parameter M. Note

[9] Note that there is a unique Nash equilibrium (for any of the 6 BCGs defined in this paper) where all players play zero.

[10] Due to an unintended typographical error, the second question of the CRT was shown to the participants as follows: *If it takes 5 machines 5 minutes to make 1 widget, how long would it take 100 machines to make 100 widgets?*. In this case, the intuitive response is not 100, as in the original version. Since we analyze players’ behavior according to the number of correct answers, and not their impulsiveness, this has not a significant impact in our results. Note that the correct answer now, 25 minutes, is a little bit more difficult to calculate. We have replicated the analysis using only questions one and three in the CRT, and we don’t find substantial differences in the results.

[11] For example, the third question is particularly interesting in our case as it requires some recursive thinking to be solved. This could be the case for the BCG and the way to think about it in a step-by-step reasoning procedure (see Coricelli and Nagel, 2009 for an extensive discussion)

[12] Only 23% (44 out of 191), 9% (17 out of 191), and 1% (2 out of 191) of the subjects scored 1, 2 and 3 on the CRT, respectively. Frederick (2005) reports that 33%, 28%, 23%, and 17% of the participants scored 0, 1, 2, and 3, respectively.

[13] A detailed analysis between Raven and CRT items shows very interesting correlations (Spearman): the correlation with item 1 (bat = ball) is 0.1625 (p-value = 0.0247), with item 2 (machines) is 0.0821 (p-value = 0.2589) and with item 3 (pads) is 0.3230 (p-value = 0.000).

[14] CRT score = 2, 3 have been pooled in a single category due to the lack of observations. Ravenś score was also transformed to a categorical variable.

[15] The choices ranged from 0 to 99 in the first game ($M = 2/3$) and in the last two games ($M = 1/5$, $M = 1/2$).

[16] It is important to remark here what the Raven test captures: *subjects’ ability to learn from immediate experience*.

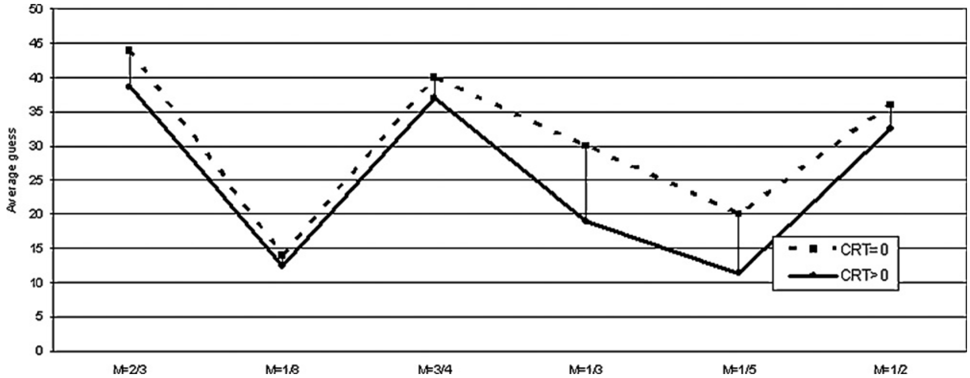

Fig. 2. Average guess by CRT.

Table 3
Estimated effects of cognitive abilities on BCG entries.

	2/3	1/8	3/4	1/3	1/5	1/2
Female	−3.09(0.39)	−1.98(0.58)	−2.45(0.53)	−5.11(0.11)	−3.84(0.30)	−4.78(0.18)
Raven	−0.37(0.20)	−0.02(0.95)	0.14(0.66)	−0.03(0.92)	0.01(0.96)	0.02(0.94)
CRT	2.52(0.31)	−2.18(0.38)	− **4.48(0.10)**	− **6.28(0.00)**	− **8.43(0.00)**	− **4.18(0.09)**
log-l	−862.10	−860.55	−878.47	−836.50	−868.57	−860.72
Wald χ^2	2.88(0.41)	0.96(0.81)	2.81(0.42)	9.16(0.03)	10.74(0.01)	3.93(0.27)

Notes: *p*-Value. Significant coefficients are shown in bold.

Table 4
BCG entries: random effects panel data.

	A	B
Female	−3.55(0.13)	−3.54(0.14)
Raven	−0.04(0.83)	−0.03(0.83)
CRT	− **3.84(0.02)**	− **3.84(0.02)**
$D_{1/8}$		−19.41(0.00)
$D_{3/4}$		0.82(0.66)
$D_{1/3}$		−14.06(0.00)
$D_{1/5}$		−17.46(0.00)
$D_{1/2}$		−7.24(0.00)
log-l	−5171.94	−5078.05
Wald χ^2	6.61(0.08)	214.05(0.00)

Notes: *p*-Value; D_M is a dummy variable for each *M*-game. Significant coefficients are shown in bold.

that, parameters are showed in the same order that players faced in the game and $M = 2/3$, the first game, is the reference. As showed before, the Raven test does not provide any insightful result, but the CRT is highly significant: subject with higher CRT score select lower entries.

Looking at dummies coefficients for each *M* (model B bottom) we find an interesting result: the smaller the parameter (the signal), the higher the response toward the equilibrium. Therefore, subjects are highly sensitive to the parameters of the game and their reactions are in line to the expected direction.

Now we focus on how people play according to Game Theory: we do analyze the number of subjects who played dominated strategies and the relation to the cognitive tests. We observe that the proportion of players that never played a dominated strategy differs according to their CRT score: 27.34% of players with CRT = 0 versus 35.94% with CRT > 0, although this difference is not significant (p-value = 0.11, proportion unilateral test).

We compute the variable *irrat* ($\in [0, 6]$) as the number of times the subject played dominated strategies, i.e., if "guess $> M * 100$".[17,18] We must emphasize that it is not the same to fail in the first game ($M = 2/3$) than in the last one ($M = 1/2$) as the last choice is assumed to be easier. In the last guess, subjects have already learned through the pure experience of the

[17] According to the definition of *irrat*, the proportion of subjects categorized as 0, 1, 2, 3, 4, and 5 was 30% (58 out of 191), 20% (38), 21% (40), 18% (34), 6% (12), and 5% (9), respectively.

[18] Note that the use of > instead of ≥ is NOT trivial. This is because when subjects guess the $= M * 100$, they are not best-responding. In any case, there are very few subjects in this extreme case.

Table 5
BCG entries: use of dominated strategies.

	Irrat.	*Exp. irrat*
Female	−0.22(0.49)	−6.11(0.12)
Raven	0.01(0.70)	0.24(0.44)
CRT	− **0.51(0.03)**	− **7.78(0.01)**
log-l	−331.92	−660.22
Wald χ^2	5.93(0.16)	8.93(0.030)

Notes: *p*-Value. Significant coefficients are shown in bold.

Table 6
Level of reasoning and abilities.

	Low level	High level
Female	−0.17(0.64)	−0.15(0.75)
Raven	−0.02(0.54)	−0.04(0.23)
CRT	− **0.44(0.09)**	**0.62(0.03)**
log-l	−99.75	−71.56
Wald χ^2	3.81(0.28)	5.88(0.12)

Notes: *p*-Value. Significant coefficients are shown in bold.

game (feedback free learning; Weber, 2003). The variable *exp_irrational* captures this idea. We define *exp_irrational* ($\in [0, 63]$) as the number of times the subject plays dominated strategies weighted by the order they were played.[19]

Table 5 below shows the results of estimating the effect of both Raven and CRT on individual rationality. We use censored Tobit regression with normal disturbances models according to the values of the dependent variable.

Once again we find that Raven does not have any explanatory power. However, CRT appears to be significant again: subjects with positive scores on the CRT are less prone to play dominated strategies. This is true for both definitions of learning.

4. Levels of reasoning

The existing literature in the BCG has shown that there are several modes in the distribution of actual plays that indicate different levels of reasoning. Following Coricelli and Nagel (2009), we compute subjects' level of reasoning and analyze whether these levels are related to cognitive abilities.

First, we compute the quadratic distance, QD_{ik}, for each i individual at each M-game from the respective k-level of reasoning[20]: $QD_{ik} = (x_{iM} - 50M^k)^2$, where x_{iM} is the entry of subject i in M-game. Round by round and individual by individual, we determine the minimum distance and the corresponding k-level of reasoning.

We categorized a player as

- *low* level if at least 4 out of 6 cases was identified as level 1 (148 subjects, 78% of the sample),
- *high* level if at least 4 out of 6 cases was identified as level 2 or higher (25, 13%),
- *random* players to those not belonging to any of previous categories (18, 9%).

We have shown above that subjects with lower score in the CRT play higher numbers and dominated strategies in the BCG more often than those with higher score in the CRT. In this section, we study if *higher CRT score is related with a higher level of reasoning in the BCG*. First, we use logit models to estimate the effect of cognitive abilities on the probability of belonging to a low-, or high-, level of reasoning (see Table 6). In line with the results shown above, subjects with a higher score in the CRT have a lower (greater) probability to belong to low (high) level category, while Raven test is not informative.

Second, and similarly to previous analysis, we run a panel data regression to explore the dynamics of level of reasoning. The dependent variable is the level of reasoning of i individual at each round. Level of reasoning varies from 0 to 4 (see footnote 19). Table 7 shows the results of a panel data analysis using random effect ordered logit regressions. Model A explores the relation between level of reasoning and cognitive abilities, whereas model B uses dummies to capture the effect of the parameter M. Raven test is not significant explaining the dependent variable, while the CRT value is positive and significant.

[19] $exp\_irrat = 2^5 \times irrat_6 + 2^4 \times irrat_5 + 2^3 \times irrat_4 + 2^2 \times irrat_3 + 2 \times irrat_2 + irrat_1$.
[20] According to Coricelli and Nagel (2009), a naïve player (level 0) will choose a number randomly. A level 1 player assumes that others play randomly (level 0) and chooses 33 (=2/3*50), giving that 50 is the average of randomly chosen numbers from 0 to 100. A level 2 player assumes that others are level 1 and best responds playing 22 (=2/3*33), and so on.

Table 7
Random effects panel data.

	A	B
Female	−0.17(0.57)	−0.29(0.35)
Raven	−0.01(0.74)	0.01(0.63)
CRT	**1.07(0.00)**	**0.89(0.00)**
$D_{1/8}$		−2.22(0.00)
$D_{3/4}$		0.69(0.01)
$D_{1/3}$		−1.25(0.00)
$D_{1/5}$		−1.40(0.00)
$D_{1/2}$		−0.71(0.02)
log-l	−773.60	−724.52
Wald χ^2	41.51(0.00)	102.87(0.00)

Notes: p-Value; D_M is a dummy variable for each M-game. Significant coefficients are shown in bold.

The estimated coefficients for the dummies in Table 7, allows us to qualify previous results regarding the same dummies in Table 4. We said before that the smaller M, the higher the probability of choosing a lower entry. Table 7 shows us that, although they select lower numbers, it is not enough to achieve a higher level of reasoning. That is, subjects report lower guesses, as the result of a lower M, but this decline does not necessarily implies a jump in level of reasoning.

5. Concluding remarks

The BCG is an intriguing game in that only a tiny fraction of people are able to solve it, but once the logic of the game is revealed, most people find the Nash equilibrium to be an obvious prediction. This paper explores if people who are able to solve the BCG have higher cognitive abilities. We measure intelligence using two complementary tests: the Raven and the CRT. Our subject pool played six (incentivized) one-shot p-beauty games without any feedback. We find that subjects with higher scores on the CRT test are more prone to play according to the Nash equilibrium. In sharp contrast, the Raven test does not provide any insight on BCG playing. Furthermore, results are confirmed when levels of reasoning instead of entries in the BCG are used.

Finally, all along the paper we do not find any gender effect. Therefore, the effect of cognitive abilities, among women and men, is not different playing the BCG.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jebo.2012.05.018.

References

Ablard, K.E., Mills, C., 1996. Evaluating abridged versions of the Raven's advanced progressive matrices for identifying students with academic talent. Journal of Psychoeducational Assessment 14 (1), 54–64.
Benjamin, D.J., Brown, S.A., Shapiro, J.M. Who is "behavioral"? Cognitive ability and anomalous preferences. Journal of the European Economics Association, forthcoming.
Brañas-Garza, P., Espinosa, M.P., Rey-Biel, P., 2011. Travelers types. Journal of Economic Behavior & Organization 78 (1–2), 25–36.
Brañas-Garza, P., Guillen, P., Lopez, R., 2008. Math skills and risk attitudes. Economics Letters 99 (2), 332–336.
Bruguier, A.J., Quartz, S.R., Bossaerts, P., 2010. Exploring the nature of 'trader intuition'. Journal of Finance 65 (5), 1703–1723.
Burnham, T.C., Cesarini, D., Johannesson, M., Lichtenstein, P., Wallace, B., 2009. Higher cognitive ability is associated with lower entries in a p-Beauty Contest. Journal of Economic Behavior & Organization 72 (1), 171–175.
Camerer, C.F., Ho, T.-H., Chong, J.-K., 2004. A cognitive hierarchy model of games. The Quarterly Journal of Economics 119 (3), 861–898.
Coricelli, G., Nagel, R., 2009. Neural correlates of depth of strategic reasoning in medial prefrontal cortex. Proceedings of the National Academy of Sciences of the United States of America 106 (23), 9163–9168.
Dohmen, T., Falk, A., Huffman, D., Sunde, U., 2010. Are risk aversion and impatience related to cognitive ability? American Economic Review 100 (3), 1238–1260.
Finucane, M.L., Gullion, C.M., 2010. Developing a tool for measuring the decision-making competence of older adults. Psychology and Aging 25 (2), 271–288.
Fischbacher, U., 2007. Z-tree: Zurich toolbox for ready-made economic experiments. Experimental Economics 10, 171–178.
Frederick, S., 2005. Cognitive reflection and decision making. Journal of Economic Perspectives 19 (4), 25–42.
Gill, D., Prowse, V., 2012. Cognitive ability and learning to play equilibrium. A level-*k* analysis, Mimeo.
Horn, J.L., Cattell, R.B., 1966. Refinement and test of the theory of fluid and crystallized general intelligences. Journal of Educational Psychology 57 (5), 253–270.
Mills, C.J., Ablard, K.E., Brody, L.E., 1993. The Raven's progressive matrices: its usefulness for identifying gifted/talented students. Roeper Review 15, 183–186.
Nagel, R., 1995. Unraveling in guessing games: an experimental study. American Economic Review 85 (5), 1313–1326.
Oechssler, J., Roider, A., Schmitz, P.W., 2009. Cognitive abilities and behavioral biases. Journal of Economic Behavior & Organization 72 (1), 147–152.
Raven, J.C., 1936. Mental tests used in genetic studies: the performance of related individuals on tests mainly educative and mainly reproductive. MSc Thesis. University of London, London.
Raven, J., Raven, J.C., Court, J.H., 2000. Standard Progressive Matrices. Oxford Psychology Press.
Toplak, M., West, R., Stanovich, K., 2011. The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. Memory and Cognition 39 (7), 1275–1289.
Weber, R.A., 2003. 'Learning' with no feedback in a competitive guessing game. Games and Economic Behavior 44 (1), 134–144.