Exp Econ (2016) 19:713–726
DOI 10.1007/s10683-015-9461-0

ORIGINAL PAPER

Who knows it is a game? On strategic awareness and cognitive ability

Dietmar Fehr[1] · Steffen Huck[1,2]

Received: 8 April 2014 / Revised: 23 July 2015 / Accepted: 27 July 2015 /
Published online: 11 August 2015

Abstract We examine strategic awareness in experimental games, that is, the question of whether subjects realize they are playing a game and thus have to form beliefs about others' actions. We conduct a beauty contest game and elicit measures of cognitive ability and beliefs about others' cognitive ability. We show that the effect of cognitive ability is highly non-linear. Subjects below a certain threshold choose numbers in the whole interval and their behavior does not correlate with beliefs about others' ability. In contrast, subjects who exceed the threshold avoid choices above 50 and react very sensitively to beliefs about the cognitive ability of others.

Keywords Cognitive ability · Beliefs · Beauty contest · Strategic sophistication · Strategic awareness

JEL Classification C7 · C9 · D0

Electronic supplementary material The online version of this article (doi:10.1007/s10683-015-9461-0) contains supplementary material, which is available to authorized users.

✉ Dietmar Fehr
dietmar.fehr@wzb.eu

Steffen Huck
steffen.huck@wzb.eu

1 WZB Berlin Social Science Center, Reichpietschufer 50, 10785 Berlin, Germany

2 University College London, Drayton House, 30 Gordon Street, London WC1H OAX, UK

1 Introduction

Many interactions are characterized by rather subtle strategic structures that individuals might easily overlook. For example, a worker in a firm might think that his wage (or a future promotion) just depends on his own effort while, in reality, it is linked to the competition between different units and their managers.[1] In order to understand such strategic ripple effects it is important to recognize the variability of individual strategic sophistication and to uncover the underlying attributes.

In this paper, we link economic behavior in a strategic situation to cognitive ability and beliefs about others' cognitive ability to explore whether subjects' behavior displays some minimum strategic sophistication. Our evidence comes from a laboratory study using Nagel's (1995) classic beauty contest game (BCG). In the standard BCG, subjects have to pick a number $a \in [0, 100]$ and the subject who is closest to a fraction p of the group average wins a prize m. For $p \in (0, 1)$ the unique Nash equilibrium is zero, which is also the only outcome surviving iterated elimination of (weakly) dominated strategies.

In the last decade BCGs have been extensively used to examine rationality and belief formation, in particular, in the context of level-k models (see, for example Nagel 1995, Ho et al. 1998, Costa-Gomes and Crawford 2006, and for a review on recent evidence on strategic thinking Crawford et al. 2013). The typical level-k model assumes that higher level players anchor their best responses on the behavior of non-strategic level-0 players and that belief formation proceeds in iterative steps. Thus choices of sophisticated players depend on their model of level-0 play, their own level of reasoning and their expectation about the level of reasoning of others. From the perspective of the vast literature on level-k thinking, we take one step back and pose the question which subjects realize that reasoning about how to play the game requires reasoning about others, i.e., who is aware of the strategic setting.

We identify subjects' awareness of the importance of reasoning about others through a short and reliable measure for cognitive ability and by eliciting subjects' beliefs about others' cognitive ability. Intuitively, cognitive ability might be an important determinant of the reasoning process and, recently, several studies related subjects' behavior in the BCG directly to their cognitive ability (e.g., Burnham et al. 2009, Branas-Garza et al. 2012, Carpenter et al. 2013, Georganas et al. 2015, and Gill and Prowse 2015) or brain activity (Coricelli and Nagel 2009). If a higher level of reasoning is associated with higher cognitive ability, then we expect that those with higher cognitive ability will take their opponents' cognitive ability into account when reasoning about their opponents' sophistication. Accordingly, we posit that only subjects with high cognitive ability condition their choices in the BCG on their beliefs about the cognitive skills of opponents. While this identification is indirect, an important advantage of using cognitive ability measures is that it does not

[1] There is some evidence that individuals do not consider other individuals behavior when making decisions. For example, individuals compete too much on easy tasks and too little on hard tasks (Moore and Cain 2007), ignore competitors' characteristics that are payoff-relevant (Camerer and Lovallo 1999) or take competitors behavior as given (Goldfarb and Xiao 2011).

conflict with behavior in the game itself as would be the case with direct belief elicitation and targeted questions.

We find a strong link between cognitive ability, beliefs about others' cognitive ability and choices in the BCG. Essentially, our data shows that subjects need to have some minimal cognitive ability to realize the importance of reasoning about others in the BCG. Subjects below a certain threshold play the BCG *as if* it was a game of luck: there is no evidence for reasoning about others and their choices appear to be randomly distributed over the whole interval.[2] In contrast, subjects above the threshold avoid choices above 50 and do reason about others. More specifically, their choices depend on their beliefs about others' cognitive ability.

Our paper complements the recent literature on cognitive ability and the BCG mentioned above. While Burnham et al. (2009) and (Carpenter et al. 2013) document that lower choices in a one-shot BCG are associated with higher cognitive ability as measured through a common IQ test, other studies find no or little evidence for such a relationship when looking at initial choices (e.g., Branas-Garza et al. 2012; Georganas et al. 2015 and Gill and Prowse 2015).[3] Using fMRI data, Coricelli and Nagel (2009) find substantial evidence that higher levels of reasoning in BCGs trigger activity in certain brain areas associated with mentalizing. Our study goes further and focuses on subjects' beliefs about the cognitive ability of their opponents. The findings of our study suggest that cognitive ability is an important ingredient for strategic thinking in novel situations and more importantly that strategic reasoning is endogenously determined by expectations about others' cognitive skills.

Two closely related studies attempt to manipulate subjects' beliefs about their opponents' sophistication without measuring cognitive ability directly. Agranov et al. (2012) show that behavior shifts to higher observed levels of reasoning, when undergraduate subjects know they are playing the BCG against experienced graduate students instead of other undergraduates. In a similar vein Alaoui and Penta (2015) show that subjects' level of reasoning in an 11–20 game varies with their knowledge about the inferred sophistication of opponents.[4]

2 This finding is in line with other recent evidence on the presence of non-strategic players in the BCG. For example, a study by Agranov, Caplin and Tergiman (2015) identifies a large proportion of $k = 0$ players (43 percent) by observing players' provisional BCG choices within a 3 min time frame. Related, Burchardi and Penczynski (2014) use communication protocols among team members playing a BCG to classify strategic and non-strategic reasoning (about 33 percent).

3 Branas-Garza, Garcia-Munoz and Gonzalez (2012) use the Raven test and the cognitive reflection test (CRT) to measure subjects' cognitive ability and relate the test scores to behavior in six BCGs with varying p. They find no effect for cognitive ability in the first two beauty contests (with $p = 2/3$ and $p = 1/8$), but in the remaining BCGs ($p = 1/5, p = 1/3, p = 1/2, p = 3/4$) they observe a negative correlation of CRT and chosen numbers. Similarly, Gill and Prowse (2015) find no relationship between Raven's test scores and choices in the first round of a repeated BCG ($p = 7/10$), but they observe that more cognitively able subjects converge more frequently to equilibrium over time. Georganas, Healy and Weber (2015) find that CRT scores are related to level-2 play in a variety of 2-person BCGs.

4 Alaoui and Penta (2015) show theoretically that their experimental results are consistent with individuals engaging in a cost-benefit analysis of applying additional rounds of reasoning. See also Strzalecki (2014) for an alternative theoretical approach in which behavior also depends on own bounded rationality and beliefs about opponents' bounds.

More generally, our findings add to an emerging literature that explores the impact of cognitive ability on economic decision making. For example, the studies of Benjamin et al. (2013), Burks et al. (2009), Dohmen et al. (2010), Frederick (2005) or (Huck and Weizsäcker 1999) demonstrate that higher cognitive ability is associated with less biased risk-taking and time discounting behavior.[5] In light of this evidence, our results emphasize the importance of cognitive ability for economic behavior in strategic contexts.

2 Setup

We have data from 240 subjects playing BCGs ($p = 2/3$) in 15 computerized sessions, which were conducted at the WZB-TU laboratory in Berlin between May 2010 and June 2013.[6] Subjects were randomly matched into groups of six and the winner's prize was 5 euros, which was equally shared in case of a tie.[7] The instructions of the game, including an illustrative example, were presented on the computer screen. Before subjects proceeded to the game they had to indicate whether they understood the game and, if necessary, they could clarify any open issues. Subjects learned the outcome of the BCG at the end of a session.

We measured subjects' cognitive ability with the cognitive reflection test (CRT), introduced by Frederick (2005). The CRT is a three-question test that builds upon the presumption that our brain uses two types of cognitive processes, a more intuitive or emotional process and a more deliberate and cognitively demanding process.[8] The test measures subjects' ability to inhibit intuitive (but wrong) responses and therefore measures one particular component of cognitive ability. The CRT score correlates with various behavioral biases, such as the conjunction fallacy or the base-rate fallacy (Oechssler et al. 2009; Hoppe and Kusterer 2011), and a higher CRT score is, for example, associated with a higher likelihood of play of Nash actions in normal-form games (e.g., Grimm and Mengel 2012).

The CRT was administered before subjects learned about and played the BCG. After the test, subjects had to estimate for each of the three questions the proportion of correct answers among the subjects in their session. This provides us for each subject with a measure for the perceived difficulty of each question as well as with a

[5] Other studies find that higher cognitive ability is, for example, related to better financial decision making (Agarwal and Mazumder 2013) and stock market participation (Christelis et al. 2010) or health insurance take up (Fang et al. 2008).

[6] The BCG was conducted alongside other experimental modules (for details and subjects characteristics, see Appendix A and Fehr 2013). Subjects were recruited through ORSEE (Greiner 2015) and the experiment was run with z-tree (Fischbacher 2007).

[7] In four sessions subjects were unintentional matched in groups of twelve and thus in total there are 36 groups in the BCG. While the average choice turned out to be similar in groups of twelve (42.3) and six (45.8), the regression analysis controls for group-size effects.

[8] The three questions are the following: A bat and a ball cost 1.10 Euro in total. The bat costs 1.00 Euro more than the ball. How much does the ball cost? If it takes 5 machines 5 min to make 5 widgets, how long would it take 100 machines to make 100 widgets? 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?

measure of how they assess others' cognitive ability. We will use this measure in our analysis to identify a subject's awareness of the strategic aspects of the decision situation she is facing. Subjects were rewarded with 1 Euro for each correct answer in the CRT and with 25 cents for each correct prediction of the proportion of a correct answer to a CRT question, i.e., they could earn a modest 75 cents if all three predictions were correct.

We complemented the CRT with a five-minute 20-question variant of the Wonderlic Personnel Test (WPT).[9] This test provides a measure for general cognitive ability and we use it mainly to check the robustness of the CRT. Therefore, we did not elicit subjects' beliefs about other subjects' WPT scores. The WPT was administered at the end of a session and subjects were rewarded with 25 Cents per correct answer.[10]

3 Results

Figure 1 shows the distribution of choices in the BCG. Choices are spread out over the whole interval and there are two modes at 50 and 33 (about 68 % of choices are below 51). The average choice in the BCG is 45.1. While some studies find average choices below 40, e.g., Nagel (1995) and Agranov et al. (2012), other studies report average choice above 40 as well, e.g., Branas-Garza et al. (2012), Burchardi and Penczynski (2014) and (Gill and Prowse 2015).[11]

About 19 percent of choices are in the range between 67 and 100, which is clear evidence for a violation of one step of iterated dominance.[12] Choices in this range are associated with lower cognitive reflection: the average CRT score of subjects choosing a number between 67 and 100 is 0.89, which is substantially and significantly lower than the average CRT score of 1.63 for subjects choosing numbers below 67 ($p < 0.01$, two-sided t test).

Figure 2 shows the relationship of BCG choices and CRT score. The figure reveals a striking non-linear relationship. As can be seen, subjects who have answered two or less questions correctly choose numbers over the whole interval as

[9] The used questions can be found in Appendix D. For general information about the test see www.wonderlic.com.

[10] In one session there was a false fire alarm during the WPT and hence the test scores are not used in the analysis. The other tasks were not affected because they were completed before the false alarm.

[11] Branas-Garza et al. (2012) report an average choice of 43.2 ($p = 2/3$ BCG) and (Gill and Prowse 2015) find an average choice of 44.2 ($p = 7/10$ BCG). Using two-player teams in a $p = 2/3$ BCG, Burchardi and Penczynski (2014) find that the suggested average choices of subjects in a team are 43.9 (the final average choice was 39.7 after communication).

[12] Other studies find similar shares. For example, Branas-Garza et al. (2012) find that 16 percent of subjects choose a number in the interval [2/3*100, 100], Burnham et al. (2009) find that about 20 percent of subjects choose a number in the interval [1/2*100, 100] and Allred et al. (2014) find that 33 percent of their subjects choose a number in [2/3*10, 10]. Relatedly, Agranov et al. (2015) show that when subjects can modify their choice as often as they want within a three-minute time frame, more than 50 percent of subjects choose at least once an action in the interval [2/3*100, 100]. In their setup, non-strategic subjects switch their choices more than five times as often as strategic subjects and the average choices of non-strategic subjects are remarkably close to 50. Similarly, we find that, in particular, the choices of subjects with a CRT score of 0 or 1 are on average close to 50.

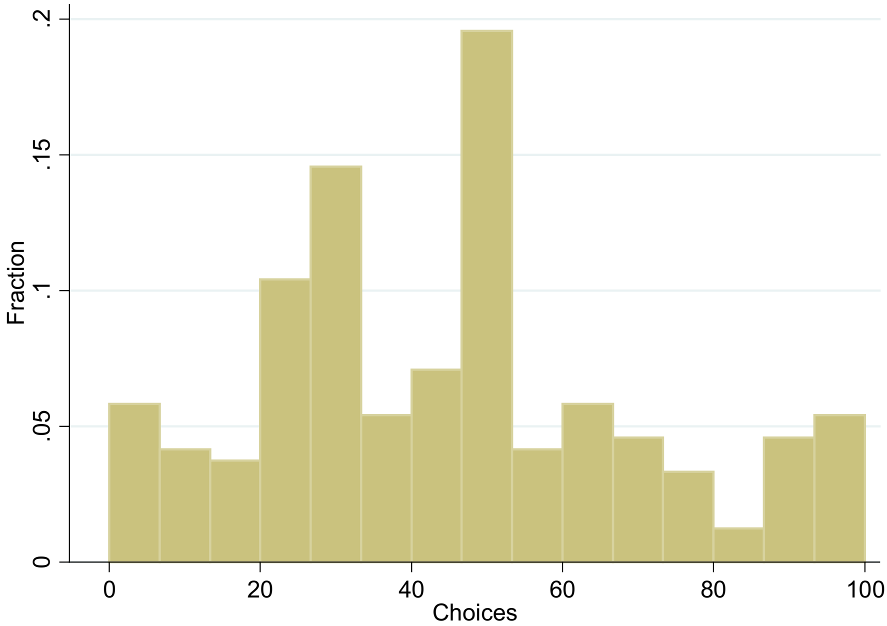

Fig. 1 Distribution of choices in the BCG

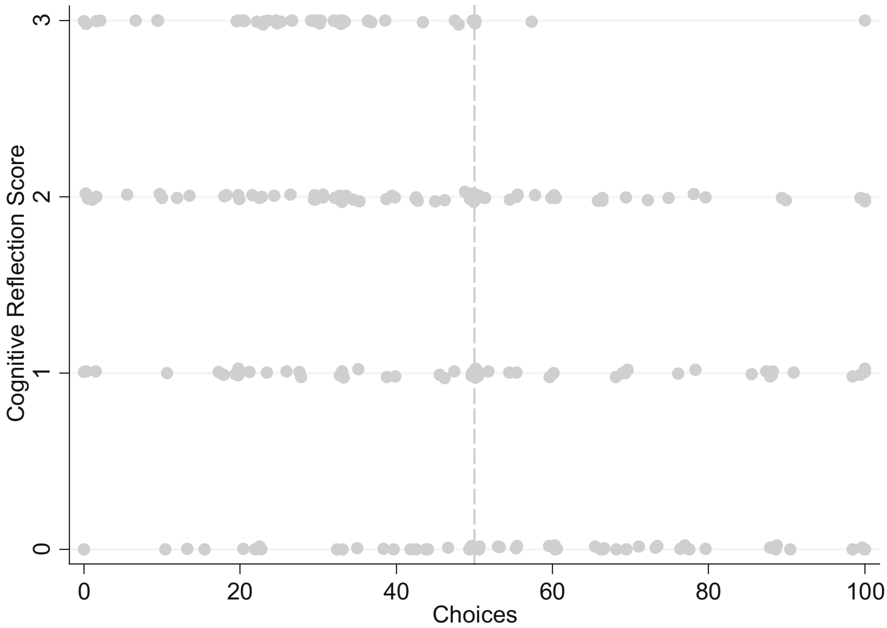

Fig. 2 Choices in the BCG and relationship with cognitive reflection (CRT score)

if they choosing randomly. In contrast, subjects who have answered all three questions correctly systematically avoid numbers above 50. We can reject that the median choice is the same for more cognitively able subjects (3 correct answers) and less cognitively able subjects (2 or less correct answers) (Mann–Whitney test, $z = 5.351$, $p < 0.01$).

Table 1 shows the distribution of correct answers for the three CRT questions as well as subjects' expectations about how many others would answer the question correctly.[13] Our subjects find the "Bat and Ball" question (henceforth BBQ) more difficult than the other two questions.[14] Only 32 percent of subjects give the correct answer to the BBQ, whereas almost twice as many give correct answers to the other two questions. Approximately two-thirds of those subjects, who answer the BBQ correctly (62 %), do answer all three questions correctly.[15] The average choice in the BCG of subjects answering all three questions correctly is 29.1, whereas the average choice is 38.3 for subjects with a correct BBQ but less than the maximum CRT score.

It is only superficially surprising that subjects estimates of others' performance are inversely related to the actual performance. Those who believe in the "intuitive" but wrong answer to the BBQ must be convinced that the answer is stunningly simple. Indeed, they expect that on average 73 % of other subjects gave the correct answer to the BBQ. In contrast, we find that subjects who got the correct answer to the BBQ have a better assessment of its difficulty. They expect that 52 percent of subjects gave the correct answer to the BBQ ($p < 0.01$, two-sided t test).

Table 2 presents estimation results from Tobit regressions that examine the relationship between choices in the BCG and cognitive ability controlling for individual characteristics such as gender, major and time enrolled. The first specification in column 1 indicates that subjects majoring in a STEM field (science, technology, engineering or mathematics) choose lower numbers than subjects from non-STEM fields. Once we include our cognitive measures, this effect is smaller. In the regression in column 2 we include a dummy variable for subjects with the maximum CRT score of three to account for the observed non-linear effect of cognitive ability. The result confirms that those subjects with the maximum score in the CRT choose significantly lower numbers in the BCG. As shown above, the questions in the CRT varied in their difficulty, and the BBQ on its own appears to be a good proxy for the maximum CRT score. In column 3 we use the BBQ as an explanatory variable instead of the maximum CRT score. The resulting coefficient is similar to the previous estimate and highly significant. This lends further support

[13] The average score in the 3-item CRT is 1.49. About 48 % of the subjects have one or less answer correct and 52 % of the subjects have 2 or 3 answers correct. Other studies find similar average scores. For instance, the mean score of the whole sample in Frederick (2005) was 1.24 (with a maximum of 2.18 and a minimum of 0.57), whereas Hoppe and Kusterer (2011) and Oechssler et al. (2009) report slightly higher scores of 1.84 and 2.05 among German students, respectively.

[14] The "Bat and Ball" question (BBQ) refers to the following question: A bat and a ball cost 1.10 Euro in total. The bat costs 1.00 Euro more than the ball. How much does the ball cost?

[15] It is also the case that subjects who answered the BBQ correctly were more likely to win the prize ($p = 0.04$, χ^2-test).

Table 1 Distribution of answers in the CRT

Question	Answer			Expected correct
	Correct (%)	Intuitive (%)	Other (%)	Answers (%)
Bat and Ball (BBQ)	32	60	8	65
Widgets	59	29	12	59
Lake	58	21	21	52

BBQ A bat and a ball cost 1.10 Euro in total. The bat costs 1.00 Euro more than the ball. How much does the ball cost? *Widgets* If it takes 5 machines 5 min to make 5 widgets, how long would it take 100 machines to make 100 widgets? *Lake* 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?

to the observation that the BBQ on its own is an important indicator for cognitive ability and subsequently for choices in the BCG.[16]

The CRT measures subjects' ability to reflect on a problem and to suppress intuitive responses as well as mathematical skills. This is only one particular aspect of cognitive ability and we want to see whether the observed relationship also holds for a more general measure of cognitive ability such as the WPT score. The correlation between the CRT score and the WPT score is about 0.47 and statistically significant ($p < 0.01$), which indicates that the two tests indeed capture a similar underlying cognitive trait. Therefore it is not surprising to see that the association between CRT score and choices in the BCG is robust to using the WPT score as an alternative measure for cognitive ability (column 4). This result also suggests that even though the CRT is arguably a coarse measure for cognitive ability (only three questions), the questions (in fact, the BBQ alone) are powerful enough to detect substantial differences in cognitive ability in the context of our experimental game.

We hypothesized in our introduction that cognitive ability is related to subjects' understanding that they have to think about others' strategic sophistication. While we have seen that those who fail on at least one of the CRT questions behave as if choosing randomly, we have not yet seen direct evidence that those who did answer all questions correctly do reason about others. For that purpose we will now utilize subjects' beliefs about others' cognitive ability. Note that eliciting beliefs about others' choices in a BCG with standard instructions would in essence be equivalent to asking for their strategic choice (albeit multiplied by p). Moreover, eliciting beliefs directly may induce strategic reasoning as it makes the strategic aspect of the

[16] That the BBQ is an important indicator for choices in the BCG is also confirmed by testing the predictive power that each of the three CRT questions provides, conditional on the other questions. That is, including a dummy variable for each correctly answered question reveals that the BBQ has by far the largest the impact on choices in the BCG and we can reject the hypothesis that all three questions are equally predictive (Wald test, $p = 0.058$, two sided).

Table 2 Regressions—cognitive ability and choices

	Dependent variable: choice in beauty-contest game			
	(1)	(2)	(3)	(4)
CRT score = 3 (d)		−19.534***		
		(3.792)		
BBQ (d)			−18.573***	
			(2.811)	
WPT score				−3.522***
				(0.690)
Female (d)	4.841	5.126	4.624	4.977*
	(3.265)	(3.165)	(3.038)	(2.775)
Major: STEM (d)	−9.625**	−7.173**	−6.188*	−6.842
	(3.717)	(3.299)	(3.178)	(4.187)
# of terms enrolled	4.091*	4.535**	4.994**	5.347**
	(2.339)	(2.243)	(2.384)	(2.285)
Instruction time	3.830	1.308	0.900	0.591
	(4.414)	(4.257)	(4.536)	(3.577)
Group size (d)	−5.595	−4.441	−6.683**	−6.775*
	(3.426)	(3.203)	(3.182)	(3.863)
Constant	27.913	40.307**	43.452**	78.389***
	(19.884)	(19.624)	(20.333)	(14.297)
N	240	240	222	240

Tobit regressions with standard errors clustered on the session level in parentheses. "CRT score 3 (d)" is a dummy variable, which equals one if a subject answered all three questions in the CRT correct. "WPT Score" is the number of correct answers in the WPT, ranging from 6 to 17. "BBQ (d)" is a dummy variable and equals 1 for a correct answer to the Bat and Ball question (BBQ). "Major: STEM (d)" is a dummy variable indicating whether a subject is enrolled in science, technology, engineering or mathematics. The number of terms enrolled denotes the time enrolled at the university in log(semester). "Instruction time" is the time taken for reading the instructions and questions in log(seconds). "Group size (d)" indicates sessions were the group size was n = 12 instead of n = 6

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

BCG more salient.[17,18] Thus we prefer to examine whether subjects' beliefs about others' cognitive ability correlates with their choice in the BCG.

In our analysis we focus on subjects' beliefs about how others will do in the BBQ. We concentrate on this question because previous studies revealed that it is

[17] A similar problem may arise by using post-experimental questions. Even though such questions cannot influence behavior in the game, they can induce biased responses because subjects may realize ex-post that the decision environment was strategic. Note that in the BCG it is not possible to distinguish cleanly between non-strategic and strategic behavior, since low numbers might be evidence for both.

[18] It also seems possible that administering the CRT before the BCG makes subjects who gave the correct answers more suspicious and raises their awareness in the subsequent BCG, because giving the correct answer often requires suppressing the intuitive, but wrong answer. Note that, in principle, changing the order of the tasks may lead to the related issue of BCG-play influencing awareness in the CRT. However, as pointed out by a referee, it seems unlikely that a potentially higher awareness due to the task order would only affect subjects with the maximum CRT score of three.

Table 3 Regressions—beliefs about cognitive ability and choices

	Dependent variable: choice in beauty-contest game					
	(1) CRT = 3	(2) CRT < 3	(3) CRT = 3	(4) CRT < 3	(5) BBQ correct	(6) BBQ wrong
Expected proportion	−15.693**	−6.333			−11.748**	−7.309
BBQ correct	(7.498)	(6.363)			(5.519)	(6.709)
Avg. expected			−12.747	−5.032		
proportion correct			(8.321)	(10.052)		
Female (d)	0.715	6.207	−0.641	6.253	−1.941	7.744*
	(5.401)	(3.762)	(5.623)	(3.809)	(3.964)	(4.386)
Major: STEM (d)	0.914	−8.584**	1.112	−8.896**	−4.001	−6.372
	(8.882)	(4.105)	(8.261)	(4.123)	(8.707)	(4.224)
# of terms enrolled	0.125	5.641**	0.103	5.441**	3.388	5.782**
	(5.908)	(2.343)	(5.647)	(2.223)	(4.964)	(2.351)
Instruction time	−9.898*	4.105	−8.986*	4.179	−2.536	1.657
	(5.072)	(5.280)	(5.168)	(5.353)	(4.595)	(6.131)
Group size (d)	−8.887**	−6.170	−8.732*	−5.844	−10.934***	−7.636**
	(3.916)	(3.900)	(4.363)	(3.826)	(3.455)	(3.857)
Constant	81.275***	31.261	76.680***	30.074	50.400**	42.945
	(23.501)	(24.479)	(23.094)	(27.005)	(20.494)	(28.834)
N	48	192	48	192	77	163

Tobit regressions with standard errors clustered on the session level in parentheses. "Expected proportion BBQ correct" is the expected proportion of correct answers to the BBQ in a session (ranging from 0 to 1) and "Average expected proportion correct" is the average expected proportion of correct answers to each CRT question. "Major: STEM (d)" is a dummy variable indicating whether a subject is enrolled in science, technology, engineering or mathematics. The number of terms enrolled denotes the time enrolled at the university in log(semester). "Instruction time" is the time taken for reading the instructions and questions in log(seconds). "Group size (d)" indicates sessions were the group size was $n = 12$ instead of $n = 6$

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

the most difficult question (Oechssler et al. 2009; Hoppe and Kusterer 2011; Meyer, Spunt and Frederick 2013) and thus it may give us the best assessment of what subjects think about other subjects' cognitive ability.[19] If our hypothesis about cognitive ability and subjects' awareness of strategic aspects is correct, we should find that beliefs about others' performance in the BBQ matter for subjects with the maximum CRT score of three but not for subjects with a lower CRT score.

The Tobit regressions in column 1 and 2 of Table 3 present our main evidence. In the first column we focus only on subjects that did achieve the maximum CRT score, whereas in column 2 we only examine subjects with a lower score. The difference between both columns is pronounced. While the BCG choices of subjects who answered all three CRT questions correctly are highly sensitive to their beliefs

[19] Relatedly, the BBQ (and more generally the CRT) is a good predictor for rational thinking, i.e., individuals' immunity to cognitive biases (see e.g., Toplak et al. 2011).

about how other subjects answered the BBQ, other subjects' choices do not correlate at all with their beliefs about others' cognitive ability.[20] We should note that this result not necessarily implies that subjects actively think about others' ability to answer the BBQ when choosing their number. It rather suggests that subjects who are perfect in the CRT not only have a better understanding of the task but are also inclined to reason about others and behave strategically.

This result is robust to using an alternative belief measure and sample split. Instead of using only a subject's expectation about correct answers to the BBQ, we can use all three belief statements of a subject. That is, we consider a subject's average expectation of correct answers to the three questions as an explanatory variable (columns 3 and 4). However, we have to keep in mind that this is only a coarse measure for beliefs about the overall cognitive ability of others and also does not account for the observed non-linearity. Nevertheless, the coefficient estimates are remarkably close to the estimates of our main specification in both subsamples. They are less precisely estimated and for subjects with the maximum score in the CRT the coefficient is at the margin of being statistically significant ($p = 0.067$, one-sided; column 3). In the last two columns of Table 3, we use an alternative sample split, which allows us to focus directly on the beliefs of those who answered the BBQ correctly. Accordingly, we split the sample into subjects with a correct answer to the BBQ and into subjects with a wrong answer to that question and focus on the estimated proportion of correct answers to the BBQ within a session (as in our main specification). Again, we find that the corresponding coefficient for "Expectation proportion BBQ correct" is significant at the five-percent level for subjects with a correct BBQ (column 5), but not for subjects who fail on the BBQ (column 6). In summary, we find that expectations about the cognitive ability of others only matter for the choices of subjects above a certain cognitive-ability threshold.

4 Concluding remarks

In this note we explored subjects' understanding that reasoning about others is a prerequisite for reasoning about how to play the game. We call this form of minimal strategic sophistication "strategic awareness". An essential requirement for strategic awareness is the formation of beliefs about others' behavior. Importantly, these beliefs need not be based on equilibrium models of other players' behavior, as in alternative models of strategic thinking beliefs can be based on non-equilibrium behavior, heuristics, etc.

We present evidence for a non-linear effect of cognitive ability and choices in the BCG. That is, choices of less cognitive able subjects appear to be randomly distributed over the whole interval and their choices are not correlated with their

[20] Gill and Prowse (2015) find a similar result in an environment that facilitates learning. Using data from a repeated BCG they find that subjects with higher cognitive ability vary their level of reasoning depending on the known cognitive ability of their opponents, whereas low ability subjects do not react at all to the cognitive ability of their opponents. However, they do not find differences between high and low ability subject for initial responses in the BCG.

beliefs about others' cognitive ability. While this is suggestive for a lack of strategic awareness, it does not preclude the possibility that they are aware of the necessity to reason about others, but behave *as if* they don't understand the strategic situation because they are, for example, not capable to form expectations about others' choices. In contrast, subjects with high cognitive ability avoid choices above 50 and their choices are correlated with their beliefs about others' cognitive ability. These findings suggest that only cognitively able subjects realize that reasoning about others is essential for playing the game.

Strategic awareness will often convey important advantages in life. For example, in retail finance markets consumers' failure to understand the financial incentives of advisors can cause serious harm (see, e.g., Inderst and Ottaviani, 2012 and Chater et al. 2010) and disclosure rules that create transparency and promote the salience of strategic situation may be an important instrument for consumer protection. There are, of course, also situations where a lack of strategic awareness can be advantageous by inducing *de facto* commitment power. Cournot experiments are a good example, where subjects that simply choose quantities somewhere in the middle of the action space and do not (best) respond to others unwittingly become Stackelberg leaders outperforming their more sophisticated competitors, see e.g., (Huck et al. 1999).

While we found evidence for strategic awareness in a specific environment, it is an open question how strategic awareness depends on context. It is, for example, not clear how different aspects of intelligence affect strategic awareness or how the salience of the strategic aspects of a situation promotes strategic awareness.[21] Our results also raise the question whether strategic awareness can be promoted, for example, by providing hints and nudges or by manipulating beliefs directly.[22] A good example for such a hint is the frequently observed sticker on big trucks, which says "If you can't see my mirrors, I can't see you", to alert car drivers that they are in the blind spot of the truck driver. More research in these directions appears fruitful for both, fundamental and policy-related applied research.

Acknowledgments We thank the Editor, Jacob Goeree, and two anonymous reviewers for thoughtful comments. We are also grateful to Terry Burnham, Brit Grosskopf, Rosemarie Nagel, Joerg Oechssler and Andrew Schotter for helpful conversations as well as seminar participants for comments. We thank David Cesarini, Pablo Branas-Garza and Teresa Garcia-Munoz for sharing their data. Financial support from the Deutsche Forschungsgemeinschaft (DFG) through the SFB 649 "Economic Risk" is gratefully acknowledged.

[21] For example, different framing of instructions may induce strategic awareness. Suggestive evidence comes from 2-person guessing games where the choice of zero is weakly dominant (see e.g., Grosskopf and Nagel 2008). Chou et al. (2009) show that, while standard descriptions of the payoff function leads to frequent violations of weak dominance, framing the game as a "battle"—you win the battle if your chosen location is higher than your opponent's—improves compliance dramatically. However, their setup leaves open whether subjects just followed the advice in the instructions, or whether the variation indeed induced subjects to reason strategically.

[22] Although we can rule out that our finding is affected by reverse causality because subjects were not aware of the BCG when stating their beliefs about others' cognitive ability, exogenous variation in beliefs would be useful to account for unobserved factors that may influence subjects' beliefs about the cognitive skills of others.

References

Agarwal, S., & Mazumder, B. (2013). Cognitive abilities and household financial decision making. *American Economic Journal: Applied Economics, 5*(1), 193–207.

Agranov, M., Caplin, A., & Tergiman, C. (2015). Naive play and the process of choice in guessing games. *Journal of the Economic Science Association* (forthcoming).

Agranov, M., Potamites, E., Schotter, A., & Tergiman, C. (2012). Beliefs and endogenous cognitive levels: An experimental study. *Games and Economic Behavior, 75*(2), 449–463.

Alaoui, L., & Penta A. (2015). Endogenous depth of reasoning. *Review of Economic Studies* (forthcoming).

Allred, S., Duffy, S., & Smith, J. (2014). *Cognitive load and strategic sophistication*. New York: Mimeo.

Benjamin, D. J., Brown, S. A., & Shapiro, J. M. (2013). Who is behavioral? Cognitive ability and anomalous preferences. *Journal of the European Economic Association, 11*(6), 1231–1255.

Branas-Garza, P., Garcia-Munoz, T., & Gonzalez, R. (2012). Cognitive effort in the beauty contest game. *Journal of Economic Behavior & Organization, 83*(2), 254–260.

Burchardi, K., & Penczynski, S. (2014). Out of your mind: Eliciting individual reasoning in one shot games. *Games and Economic Behavior, 84*, 39–57.

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

Burnham, T. C., Cesarini, D., Johannesson, M., Lichtenstein, P., & 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., & Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review, 89*(1), 306–318.

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

Chater, N., Huck, S., & Inderst, R. (2010). Consumer decision making in retail financial services. Report prepared for SANCO/EC.

Chou, E., Margaret, M., Rosemarie, N., & Plott, C. (2009). The control of game form recognition in experiments: Understanding dominant strategy failures in a simple two person "guessing" game. *Experimental Economics, 12*(2), 159–179.

Christelis, D., Tullio, J., & Padula, M. (2010). Cognitive abilities and portfolio choice. *European Economic Review, 54*(1), 18–38.

Coricelli, G., & Nagel, R. (2009). Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proceedings of the National Academy of Sciences: Economic Sciences, 106*(23), 9163–9168.

Costa-Gomes, M. A., & Crawford, V. P. (2006). Cognition and behavior in two-person guessing games: An experimental study. *American Economic Review, 96*(5), 1737–1768.

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

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.

Fang, H., Keane, M. P., & Silverman, D. (2008). Sources of advantageous selection: Evidence from the medigap insurance market. *Journal of Political Economy, 116*(2), 303–350.

Fehr, D. (2013). *Foregone costly communication and coordination failure*. New York: Mimeo.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics, 10*(2), 171–178.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25–42.

Georganas, S., Healy, P., & Weber, R. (2015). On the persistence of strategic sophistication. *Journal of Economic Theory* (forthcoming).

Gill, D., & Prowse, V. (2015). Cognitive ability, character skills, and learning to play equilibrium: A level-*k* analysis. *Journal of Political Economy* (forthcoming).

Goldfarb, A., & Xiao, M. (2011). Who thinks about the competition? Managerial ability and strategic entry in US local telephone markets. *American Economic Review, 101*(5), 3130–3161.

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

Grimm, V., & Mengel, F. (2012). An experiment on learning in a multiple games environment. *Journal of Economic Theory, 147*(6), 2220–2259.
Grosskopf, B., & Nagel, R. (2008). The two-person beauty contest. *Games and Economic Behavior, 62*(1), 93–99.
Ho, T.-H., Camerer, C., & Weigelt, K. (1998). Iterated dominance and iterated best response in experimental p-beauty contests. *American Economic Review, 88*(4), 947–969.
Hoppe, E., & Kusterer, D. J. (2011). Behavioral biases and cognitive reflection. *Economics Letters, 110*(2), 97–100.
Huck, S., Normann, H.-T., & Oechssler, J. (1999). Learning in cournot oligopoly: An experiment. *Economic Journal, 109*(454), C80–C95.
Huck, S., & Weizsäcker, G. (1999). Risk, complexity, and deviations from expected-value maximization. *Journal of Economic Psychology, 20*(6), 699–715.
Inderst, R., & Ottaviani, M. (2012). How (not) to pay for advice: A framework for consumer financial protection. *Journal of Financial Economics, 105*(2), 393–411.
Meyer, A., Spunt, B., & Frederick, S. (2013). *The bat and ball problem*. New York: Mimeo.
Moore, D. A., & Cain, D. (2007). Overconfidence and underconfidence: When and why people underestimate (and overestimate) the competition. *Organizational Behavior and Human Decision Processes, 103*(2), 197–213.
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.
Strzalecki, T. (2014). Depth of reasoning and higher order beliefs. *Journal of Economic Behavior & Organization, 108*, 108–122.
Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Memory and Cognition*, *39*(7), 1275–1289.