Cognitive effort in the Beauty Contest Game[*]

Pablo Brañas-Garza & Teresa García-Muñoz
GLOBE: Universidad de Granada, Spain

Roberto Hernán[†]
Economic Science Institute, Chapman University, USA

September 19, 2011

Abstract

This paper analyzes cognitive effort in 6 different one-shot 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.

keywords: Beauty Contest Game, Raven, Cognitive Reflection Test

1 Introduction

Recent papers connect individuals' cognitive abilities with performance in different games through different tests (see for instance Burnham et al., 2009; Oechssler, Roider and Schmitz, 2009; Brañas–Garza, Espinosa and Rey-Biel, 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

---

[†]Corresponding author: Economic Science Institute, Chapman University, roberto.hernangonzalez@gmail.com

gains. Benjamin, Brown and Shapiro (2006) 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, Guillen and López (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, Guillen and Lopez (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, Ho and Chong, 2004). A higher level of reasoning indicates higher strategic behavior by subjects and the belief that rivals are also more strategic.

Burnham et al. (2009) investigated the relationship between cognitive abilities and choices in a Beauty Contest Game (BCG). They found that individuals with higher scores on the cognitive test choose numbers closer to the Nash equilibrium in the one-shot BCG. 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[1]. 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, Quartz and Bossaerts (2010) find that skill in predicting price changes in markets with insider correlates with scores

---

[1] Burnham et al. (2009) study a BCG with a parameter of p=1/2. Therefore, numbers higher than 50 are dominated by 50.

[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).

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].

We analyze the Raven test and the CRT as they have appealing characteristics for playing the BCG. Raven's Progressive Matrices test (Raven, Raven and Court, 2000) measures visual reasoning and analytic intelligence, the capacity to learn from immediate experience with the problem without rely ing on previous knowledge, and mathematical reasoning (Mills, Ablard and Brody, 1993; Ablard and Mills, 1996). 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 a good indicator of mathematical skills.

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).

2.1   Beauty Contest Game

The Beauty Contest game[4] 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 $3/4$.

---

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

[4]The original instructions are in Spanish. The instructions were provided by Rosemarie Nagel.

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). All the subjects played the different versions of the game in the same order:

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: $M = 1/2$

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 allows us to distinguish between players who play random numbers and those thinking about their best strategy [5].

2.2 Raven & CRT tests

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 Figure 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.

---

[5]We did not organize the subject pool into smaller groups with different orders to keep the large size of the pool.

Figure 1: An example of one test item in the Raven's test
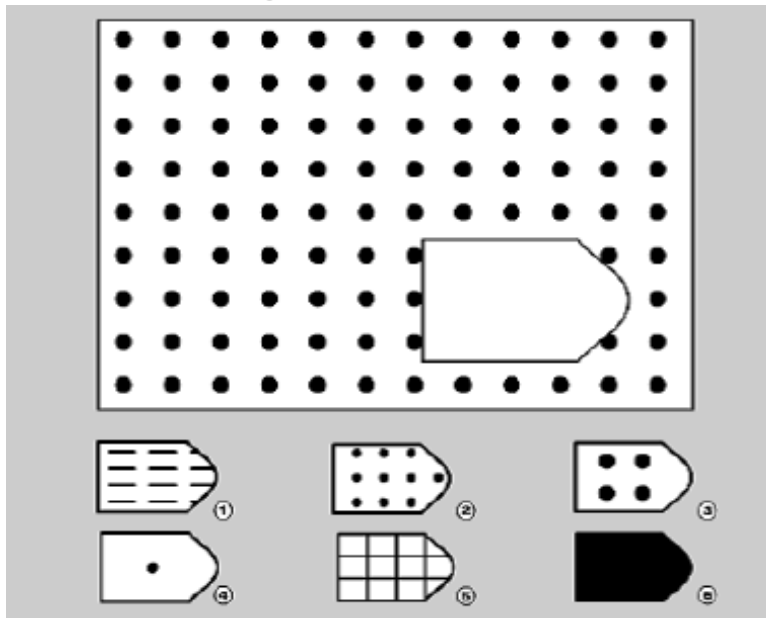
The final score is a measure of ability for abstract analytic reasoning and fluid intelligence, that is, an ability that does not rely on knowledge or skill acquired from experience as opposed to crystallized intelligence (see Horn and Catell, 1966). Following Burnham et al. (2009), we expected to find a negative[6] 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?[7]

---

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

[7]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.

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[8]. In this sense, the CRT can be viewed as a combination of cognitive capacity and the disposition for judgement and decision making (Finucane and Guillon, 2010; Toplak, West and Stanovich, 2011). Toplak, West and Stanovich (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 results are, on average, not so good (128 individuals, 67% of the sample pool, did not provide any right answers) compared to those shown in Frederick (2005)[9].

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 $\rho = 0.29; p - value < 0.000$).

---

[8]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)

[9]Only 23% (44 out of 191) and 9% (17 out of 191) of the subjects scored 1 and 2 on the CRT, respectively. None of the subjects responded correctly to the three questions. Frederick (2005) reports that 33%, 28%, 23%, and 17% of the participants scored 0, 1, 2, and 3, respectively.

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[10]. Figure 2 shows the mean values of the subjects' choices, which were classified for their score on the CRT (CRT=0 vs. CRT>0). It is easy to see that subjects' choices are related to their performance on the CRT.
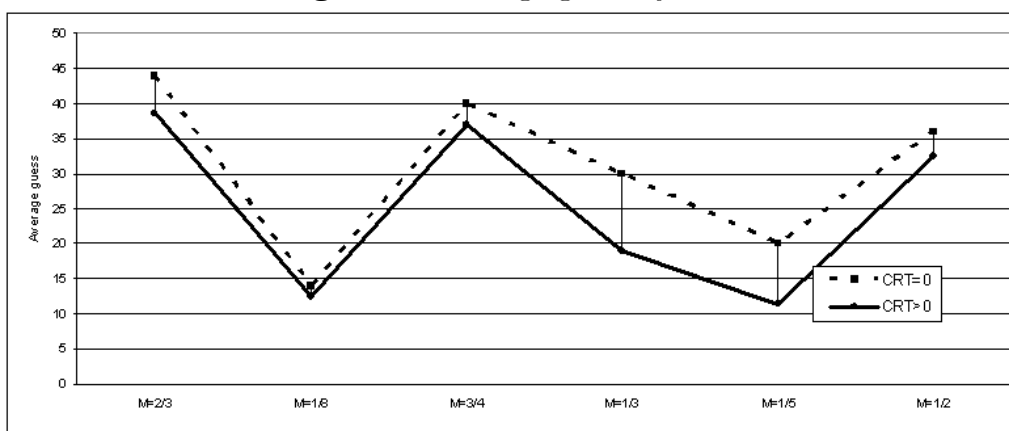
Figure 2: Average guess by CRT

Table 1 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[11].

---

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

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

7

Table 1: Estimated effects of cognitive abilities

	2/3	*1/8*	*3/4*	*1/3*	*1/5*	*1/2*
female	−3.09	−1.98	−2.45	−5.11	−3.84	−4.78
	(0.39)	(0.58)	(0.53)	(0.11)	(0.30)	(0.18)
Raven	−0.37	−0.02	0.14	−0.03	0.01	0.02
	(0.20)	(0.95)	(0.66)	(0.92)	(0.96)	(0.94)
CRT	2.52	−2.18	−4.48	−6.28	−8.43	−4.18
	(0.31)	(0.38)	(**0.10**)	(**0.00**)	(**0.00**)	(**0.09**)

*(p-value)

We study also how subjects play across games. First, we analyze the number of players 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 $> M * 100$"[12]. 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 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[13].

Table 2 below shows the results of estimating the effect of both Raven and CRT on rationality. We use censored Tobit regression with normal disturbances models according to the values of the dependent variable.

---

[12]Note that the use of $>$ instead of $\geq$ 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.

[13]$\exp\_irrat = 2^5 \times irrat_6 + 2^4 \times irrat_5 + 2^3 \times irrat_4 + 2^2 \times irrat_3 + 2 \times irrat_2 + irrat_1$

Table 2: Learning across tasks

	Irrat.	Exp. Irrat
female	−0.22	−6.11
	(0.49)	(0.12)
Raven	0.01	0.24
	(0.70)	(0.44)
CRT	−0.51	−7.78
	(**0.03**)	(**0.01**)

*(p-value)

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 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.

References

[1] Ablard, K. E. and C. Mills, 1996. Evaluating Abridged Versions of the Ravens Advanced Progressive Matrices for Identifying Students with Academic Talent. Journal of Psychoeducational Assessment 14(1): 54-64

[2] Benjamin, D. J., S. A. Brown and J. M. Shapiro, 2006. Who is "Behavioral"? Cognitive Ability and Anomalous Preferences. Levine's Working Paper Archive 122247000000001334, David K. Levine.

[3] Brañas-Garza, P., Espinosa, M. P. and P. Rey-Biel, 2011. Travelers Types. Journal of Economic Behavior & Organization 78(1-2): 25-36.

[4] Brañas-Garza, P., P. Guillen and R. Lopez, 2008. Math Skills and Risk Attitudes. Economics Letters 99(2): 332-336.

[5] Bruguier, A. J., S. R. Quartz, P. Bossaerts, 2010. Exploring the Nature of 'Trader Intuition', Journal of Finance 65 (5): 1703-1723.

[6] Burnham, T. C., D. Cesarini, M. Johannesson, P. Lichtenstein and B. Wallace, 2009. Higher Cognitive Ability is Associated with Lower Entries in a p-Beauty Contest. Journal of Economic Behavior & Organization 72(1): 171-175.

[7] Camerer, C. F., T.-H Ho and J.-K. Chong, 2004. A Cognitive Hierarchy Model of Games. The Quarterly Journal of Economics 119(3): 861-898.

[8] Coricelli, G. and R. Nagel, 2009. Neural Correlates of Depth of Strategic Reasoning in Medial Prefrontal Cortex. Proceedings of the National Academy of Sciences 106(23): 9163-9168.

[9] Dohmen, T., A. Falk, D. Huffman and U. Sunde, 2010. Are Risk Aversion and Impatience Related to Cognitive Ability? American Economic Review 100(3): 1238-60.

[10] Finucane, M. L., C. M. Gullion, 2010. Developing a Tool for Measuring the Decision-Making Competence of Older Adults, Psychology and Aging 25(2): 271–288

[11] Fischbacher, U. , 2007. Z-tree: Zurich toolbox for ready-made economic experiments. Experimental Economics, 10:171Ð178.

[12] Frederick, S., 2005. Cognitive Reflection and Decision Making. Journal of Economic Perspectives 19(4): 25-42.

[13] Horn, J. L . and R. B. Cattell, 1966. Refinement and Test of the Theory of Fluid and Crystallized General Intelligences. Journal of Educational Psychology 57(5): 253-270.

[14] Mills, C. J., K. E. Ablard, and L. E. Brody, 1993. The Raven's Progressive Matrices: Its usefulness for identifying gifted/talented students. Roeper Review 15: 183-186.

[15] Nagel, R., 1995. Unraveling in Guessing Games: An Experimental Study. American Economic Review 85(5): 1313-1326.

[16] Oechssler, J., A. Roider, P. W. Schmitz, 2009. Cognitive abilities and behavioral biases. Journal of Economic Behavior & Organization 72 (1): 147–152.

[17] Raven, J. C., 1936. Mental Tests Used in Genetic Studies: The Performance of Related Individuals on Tests Mainly Educative and Mainly Reproductive. *MSc Thesis*, London: University of London.

[18] Raven, J., J. C. Raven and J. H. Court, 2000. Standard Progressive Matrices. Oxford Psychology Press

[19] Toplak, M., R. West, K. Stanovich, 2011. The Cognitive Reflection Test as a Predictor of Performance on Heuristics-and-Biases Tasks, Memory and Cognition, forthcoming.

[20] Weber, R. A., 2003. 'Learning' with No Feedback in a Competitive Guessing Game. Games and Economic Behavior 44 (1): 134–144.

Economic Science Institute Working Papers

2011

11-07 Grether, D., Porter, D., and Shum, M. Intimidation or Impatience? Jump Bidding in On-line Ascending Automobile Auctions.

11-06 Rietz, T., Schniter, E., Sheremeta, R., and Shields, T. Trust, Reciprocity and Rules.

11-05 Corgnet, B., Hernan-Gonzalez, R., and Rassenti, S. Real Effort, Real Leisure and Real-time Supervision: Incentives and Peer Pressure in Virtual Organizations.

11-04 Corgnet, B. and Hernán-González R. Don't Ask Me If You Will Not Listen: The Dilemma of Participative Decision Making.

11-03 Rietz, T., Sheremeta, R., Shields, T., Smith, V. Transparency, Efficiency and the Distribution of Economic Welfare in Pass-Through Investment Trust Games.

11-02 Corgnet, B., Kujal, P. and Porter, D. The Effect of Reliability, Content and Timing of Public Announcements on Asset Trading Behavior.

11-01 Corgnet, B., Kujal, P. and Porter, D. Reaction to Public Information in Markets: How Much Does Ambiguity Matter?

2010

10-22 Mago, S., Sheremeta, R., and Yates, A. Best-of-Three Contests: Experimental Evidence.

10-21 Kimbrough, E. and Sheremeta, R. Make Him an Offer He Can't Refuse: Avoiding Conflicts Through Side Payments.

10-20 Savikhim, A. and Sheremeta, R. Visibility of Contributions and Cost of Inflation: An Experiment on Public Goods.

10-19 Sheremeta, R. and Shields, T. Do Investors Trust or Simply Gamble?

10-18 Deck, C. and Sheremeta, R. Fight or Flight? Defending Against Sequential Attacks in the Game of Siege.

10-17 Deck, C., Lin, S. and Porter, D. Affecting Policy by Manipulating Prediction Markets: Experimental Evidence.

10-16 Deck, C. and Kimbrough, E. Can Markets Save Lives? An Experimental Investigation of a Market for Organ Donations.

10-15 Deck, C., Lee, J. and Reyes, J. Personality and the Consistency of Risk Taking Behavior: Experimental Evidence.

10-14 Deck, C. and Nikiforakis, N. Perfect and Imperfect Real-Time Monitoring in a Minimum-Effort Game.

10-13 Deck, C. and Gu, J. Price Increasing Competition? Experimental Evidence.

10-12 Kovenock, D., Roberson, B.,and Sheremeta, R. The Attack and Defense of Weakest-Link Networks.

10-11 Wilson, B., Jaworski, T., Schurter, K. and Smyth, A. An Experimental Economic History of Whalers' Rules of Capture.

10-10 DeScioli, P. and Wilson, B. Mine and Thine: The Territorial Foundations of Human Property.

10-09 Cason, T., Masters, W. and Sheremeta, R. Entry into Winner-Take-All and Proportional-Prize Contests: An Experimental Study.

10-08 Savikhin, A. and Sheremeta, R. Simultaneous Decision-Making in Competitive and Cooperative Environments.

10-07 Chowdhury, S. and Sheremeta, R. A generalized Tullock contest.

10-06 Chowdhury, S. and Sheremeta, R. The Equivalence of Contests.

10-05 Shields, T. Do Analysts Tell the Truth? Do Shareholders Listen? An Experimental Study of Analysts' Forecasts and Shareholder Reaction.

10-04 Lin, S. and Rassenti, S. Are Under- and Over-reaction the Same Matter? A Price Inertia based Account.

10-03 Lin, S. Gradual Information Diffusion and Asset Price Momentum.

10-02 Gjerstad, S. and Smith, V. Household expenditure cycles and economic cycles, 1920 – 2010.

10-01 Dickhaut, J., Lin, S., Porter, D. and Smith, V. Durability, Re-trading and Market Performance.

2009

09-11 Hazlett, T., Porter, D., Smith, V. Radio Spectrum and the Disruptive Clarity OF Ronald Coase.

09-10 Sheremeta, R. Expenditures and Information Disclosure in Two-Stage Political Contests.

09-09 Sheremeta, R. and Zhang, J. Can Groups Solve the Problem of Over-Bidding in Contests?

09-08 Sheremeta, R. and Zhang, J. Multi-Level Trust Game with "Insider" Communication.

09-07 Price, C. and Sheremeta, R. Endowment Effects in Contests.

09-06 Cason, T., Savikhin, A. and Sheremeta, R. Cooperation Spillovers in Coordination Games.

09-05 Sheremeta, R. Contest Design: An Experimental Investigation.

09-04 Sheremeta, R. Experimental Comparison of Multi-Stage and One-Stage Contests.

09-03 Smith, A., Skarbek, D., and Wilson, B. Anarchy, Groups, and Conflict: An Experiment on the Emergence of Protective Associations.

09-02 Jaworski, T. and Wilson, B. Go West Young Man: Self-selection and Endogenous Property Rights.

09-01 Gjerstad, S. Housing Market Price Tier Movements in an Expansion and Collapse.

2008

08-10 Dickhaut, J., Houser, D., Aimone, J., Tila, D. and Johnson, C. High Stakes Behavior with Low Payoffs: Inducing Preferences with Holt-Laury Gambles.

08-09 Stecher, J., Shields, T. and Dickhaut, J. Generating Ambiguity in the Laboratory.

08-08 Stecher, J., Lunawat, R., Pronin, K. and Dickhaut, J. Decision Making and Trade without Probabilities.

08-07 Dickhaut, J., Lungu, O., Smith, V., Xin, B. and Rustichini, A. A Neuronal Mechanism of Choice.

08-06 Anctil, R., Dickhaut, J., Johnson, K., and Kanodia, C. Does Information Transparency Decrease Coordination Failure?

08-05 Tila, D. and Porter, D. Group Prediction in Information Markets With and Without Trading Information and Price Manipulation Incentives.

08-04 Caginalp, G., Hao, L., Porter, D. and Smith, V. Asset Market Reactions to News: An Experimental Study.

08-03 Thomas, C. and Wilson, B. Horizontal Product Differentiation in Auctions and Multilateral Negotiations.

08-02 Oprea, R., Wilson, B. and Zillante, A. War of Attrition: Evidence from a Laboratory Experiment on Market Exit.

08-01 Oprea, R., Porter, D., Hibbert, C., Hanson, R. and Tila, D. Can Manipulators Mislead Prediction Market Observers?