Elsevier Editorial System(tm) for Journal of Economic Behavior & Organization
Manuscript Draft

Manuscript Number:

Title: Three Very Simple Games and What It Takes to Solve Them

Article Type: Full Length Article

Keywords: cognition, bounded rationality, beliefs, guessing games, experiment

Corresponding Author: Dr Ondrej Rydval,

Corresponding Author's Institution:

First Author: Ondrej Rydval

Order of Authors: Ondrej Rydval; Andreas Ortmann; Michal Ostatnicky

Abstract: We study experimentally the nature of dominance violations in three minimalist dominance-solvable guessing games. Only about a third of our subjects report reasoning consistent with dominance; they all make dominant choices and almost all expect others to do so. Nearly two-thirds of subjects report reasoning inconsistent with dominance, yet a quarter of them actually make dominant choices and half of those expect others to do so. Reasoning errors are more likely for subjects with lower working memory, intrinsic motivation and premeditation attitude. Dominance-incompatible reasoning arises mainly from subjects misrepresenting the strategic nature (payoff structure) of the guessing games.

Three Very Simple Games and What It Takes to Solve Them†

Ondrej Rydval[a], Andreas Ortmann[b] and Michal Ostatnicky[c]

[a] Max Planck Institute of Economics, Kahlaische Strasse 10, D-07745 Jena, Germany; and CERGE-EI, Charles University Prague and Academy of Sciences of the Czech Republic, Politickych veznu 7, Prague 11121, Czech Republic. E-mail: rydval@econ.mpg.de
[b] CERGE-EI, Charles University Prague and Academy of Sciences of the Czech Republic, Politickych veznu 7, Prague 11121, Czech Republic. E-mail: andreas.ortmann@cerge-ei.cz
[c] CERGE-EI, Charles University Prague and Academy of Sciences of the Czech Republic, Politickych veznu 7, Prague 11121, Czech Republic. E-mail: michal.ostatnicky@centrum.cz

Corresponding author: Ondrej Rydval, CERGE-EI, Charles University Prague and Academy of Sciences of the Czech Republic, Politickych veznu 7, Prague 11121, Czech Republic. Tel: +420-224-005-111, Fax: +420-224-227-143, E-mail addresses: rydval@cerge-ei.cz, rydval@econ.mpg.de

Abstract

We study experimentally the nature of dominance violations in three minimalist dominance-solvable guessing games. Only about a third of our subjects report reasoning consistent with dominance; they all make dominant choices and almost all expect others to do so. Nearly two-thirds of subjects report reasoning inconsistent with dominance, yet a quarter of them actually make dominant choices and half of those expect others to do so. Reasoning errors are more likely for subjects with lower working memory, intrinsic motivation and premeditation attitude. Dominance-incompatible reasoning arises mainly from subjects misrepresenting the strategic nature (payoff structure) of the guessing games.

Keywords: cognition, bounded rationality, beliefs, guessing games, experiment
JEL classification: C72, C92, D83

† This research project was supported by a grant from Bank Austria and by CAPER grant No. LC542 of the Ministry of Education of the Czech Republic implemented at CERGE-EI (a joint workplace of the Center for Economic Research and Graduate Education, Charles University, and the Economics Institute of the Academy of Sciences of the Czech Republic). We thank Vince Crawford, Jan Kmenta and Laura Mentz for helpful comments; Vitezslav Babicky, Marian Krajc and Jana Richmanova for helping us conduct the experiment; Randall Engle and Richard Heitz of the Attention and Working Memory Lab, Georgia Institute of Technology for providing us with their "automated working memory span" tests; and Brian MacWhinney of the Department of Psychology, Carnegie Mellon University for providing us with his "audio digit span" test. All remaining errors are our own.

Figure 1

Game 2p2n: 2 players, 2 numbers

		Player 2	
		0	1
Player 1	0	M/2,M/2	M,0
	1	0,M	M/2,M/2

Game 2p3n: 2 players, 3 numbers

		Player 2		
		0	1	2
Player 1	0	M/2,M/2	M,0	M,0
	1	0,M	M/2,M/2	M,0
	2	0,M	0,M	M/2,M/2

Game 3p2n: 3 players, 2 numbers

Player 3's choice = 0

		Player 2	
		0	1
Player 1	0	M/3,M/3,M/3	M/2,0,M/2
	1	0,M/2,M/2	0,0,M

Player 3's choice = 1

		Player 2	
		0	1
Player 1	0	M/2,M/2,0	M,0,0
	1	0,M,0	M/3,M/3,M/3

Table 1

Session	1	2	3	4	5	6	7	8
Game	3p2n	2p2n	2p3n	3p2n	2p2n	2p3n	2p3n	3p2n
# participants	15	14	14	12	14	14	13	16

Table 2

	Total	Class A	Class A/B	Class A/C	Class B	Class B/C	Class C
All subjects	**112**	**66**	**3**	**3**	**3**	**7**	**30**
Choice=0	**62**	17	2	3	3	7	30
Choice=1	**50**	49	1	0	0	0	0
Belief=0	**49**	12	1	2	3	4	27
Belief=1	**63**	54	2	1	0	3	3
Choice=0 & Belief=0	**46**	9	1	2	3	4	27
Choice=0 & Belief=1	**16**	8	1	1	0	3	3
Choice=1 & Belief=0	**3**	3	0	0	0	0	0
Choice=1 & Belief=1	**47**	46	1	0	0	0	0

Table 3

	Total	Class A	Class A/B	Class A/C	Class B	Class B/C	Class C
Game 2p2n (28 subj.)	**100**	**57**	**4**	**0**	**4**	**11**	**25**
Choice=0	57	14	4	0	4	11	25
Choice=1	43	43	0	0	0	0	0
Game 2p3n (41 subj.)	**100**	**76**	**2**	**5**	**2**	**2**	**12**
Choice=0	39	17	0	5	2	2	12
Choice=1	61	59	2	0	0	0	0
Game 3p2n (43 subj.)	**100**	**44**	**2**	**2**	**2**	**7**	**42**
Choice=0	70	14	2	2	2	7	42
Choice=1	30	30	0	0	0	0	0

Table 4

REGRESSOR	*Model 1*			*Model 2*	*Model 3*
	marg. eff. (std. err.)	**marg. eff. (std. err.)**	**marg. eff. (std. err.)**	**marg. eff. (std. err.)**	**marg. eff. (std. err.)**
game 2p2n	-0.303** (0.133)	0.043* (0.022)	0.261** (0.123)	0.334** (0.136)	-0.241** (0.119)
game 3p2n	-0.346*** (0.114)	0.055** (0.026)	0.291*** (0.104)	0.354*** (0.120)	-0.320*** (0.108)
working memory	-0.084* (0.049)	0.018 (0.013)	0.066* (0.038)	0.106** (0.054)	-0.089* (0.052)
need for cognition	-0.108** (0.051)	0.023 (0.015)	0.085** (0.039)	0.115** (0.054)	-0.044 (0.054)
premeditation	-0.118** (0.046)	0.025** (0.012)	0.093** (0.038)	0.129** (0.056)	-0.010* (0.056)
Joint significance		***		***	
Number of subjects	69	10	30	109	109
	LR chi-square(5)=5.37, *p* =0.372			**% correctly predicted**	
	Hausman chi-square(5)=2.07, *p* =0.840			74.31	64.22

Blinded Manuscript (NO Author Details)
Click here to view linked References

Three Very Simple Games and What It Takes to Solve Them

Abstract

We study experimentally the nature of dominance violations in three minimalist dominance-solvable guessing games. Only about a third of our subjects report reasoning consistent with dominance; they all make dominant choices and almost all expect others to do so. Nearly two-thirds of subjects report reasoning inconsistent with dominance, yet a quarter of them actually make dominant choices and half of those expect others to do so. Reasoning errors are more likely for subjects with lower working memory, intrinsic motivation and premeditation attitude. Dominance-incompatible reasoning arises mainly from subjects misrepresenting the strategic nature (payoff structure) of the guessing games.

Keywords: cognition, bounded rationality, beliefs, guessing games, experiment
JEL classification: C72, C92, D83

1. Introduction

Experimental studies extensively document deviations of initial responses from equilibrium predictions in iterated-dominance-solvable games, including matrix and other normal-form games, extensive-form bargaining games and guessing games (see Costa-Gomes, Crawford and Broseta, 2001, and Costa-Gomes and Crawford, 2006, for overviews). Non-equilibrium behavior is typically attributed to subjects' non-equilibrium beliefs about others' irrationality rather than their own irrationality. Especially Costa-Gomes and Crawford (2006, hereafter CGC), through joint analysis of initial responses and information search patterns in iterated-dominance-solvable guessing games, convincingly show that many subjects' deviations from equilibrium "can be confidently attributed to non-equilibrium beliefs rather than irrationality, risk aversion, altruism, spite, or confusion." (p. 1740) CGC conclude that the findings "affirm subjects' rationality and ability to comprehend complex games and reason about others' responses to them..." (p. 1767).

However, parallel evidence on individual rationality from simpler, dominance-solvable games seems much less conclusive. In Grosskopf and Nagel's (2007a) two-player dominance-solvable guessing game, 90% of subjects violate simple dominance. Moreover, Devetag and Warglien (2007) show that nearly a quarter of their subjects cannot even correctly represent the relational structure of preferences in a two-player game similar to dominance-solvable guessing games. On the other hand, in Bone et al.'s (2006) extensive-form game against nature, only 5% of subjects violate simple dominance, which suggests that most people are in principle capable of applying dominance when it is transparent.[1]

Similar to Charness and Levin (2007), who simplify common value auctions to study the origin of the winner's curse, we examine the nature of dominance violations in three "minimalist" dominance-solvable guessing games featuring two or three players choosing among two or three strategies. Also called beauty contest games, guessing games are *ex ante* well-suited for studying individual rationality bounds without the potentially confounding effects of other-regarding and risk preferences. Guessing games of the dominance-solvable nature have the additional appeal of making a player's optimal choice independent of her beliefs about others' choices (and hence others' rationality).

[1] We refer to the second decision node of Bone et al.'s game where almost all of 152 subjects choose a four-payoff distribution that first-order stochastically dominates another such distribution, which is far above random choice (50%). However, the whole game is a two-stage game, and only about a third of the subjects detect dominance at the first (prior) decision node (using, for example, backward induction or the strategy method).

To better understand the decision-making errors our subjects commit, we ask them to report (alongside making choices) their detailed reasoning underlying the choices, and also to state their beliefs about their anonymous partners' choices (i.e., about others' rationality). We then study how the reported reasoning – classified according to dominance-compatibility by two independent examiners – translates into the subjects' stated choices and beliefs. Following the lead from psychology (e.g., Simon, 1978 and 1989; Stanovich and West, 2000) and recently experimental economics (e.g., Ballinger et al., 2007; Rydval, 2007), we also examine how subjects' reasoning classes and choices relate to their measured cognitive abilities and personality traits.

Only about a third of our subjects reason in line with dominance; they all make dominant choices and almost all expect others to do so. By contrast, nearly two-thirds of subjects report reasoning processes incompatible with dominance, yet a quarter of them actually make dominant choices and half of those expect others to do so. Reasoning errors are more likely for subjects with lower ability to maintain and allocate attention, as measured by working memory, and for subjects with lower intrinsic motivation and premeditation attitude. We further find that dominance-incompatible reasoning arises mainly from subjects misrepresenting (to themselves) the strategic nature (payoff structure) of the guessing games.

2. The guessing games

We study behavior in three symmetric dominance-solvable guessing games depicted in normal-form representation in Figure 1. A pair or a triplet of players simultaneously choose (or guess) among two (0, 1) or three (0, 1, 2) numbers. A fixed monetary prize, M, is won by the player whose choice is closest to one-half of the pair's or triplet's average choice; multiple winners divide the prize equally. Under complete information – an assumption justified by publicly announcing the games' structure – our games have a unique equilibrium in which all players choose 0. Games 2p2n and 3p2n are strict-dominance-solvable, i.e., choosing 0 yields a strictly higher payoff compared to choosing 1, for any choice(s) of the other player(s). Game 2p3n is weak-dominance-solvable, i.e., choosing 0 yields a higher or equal payoff compared to choosing 1 or 2, for any choice of the other player.

[Figure 1 about here]

Previous studies predominantly focus on iterated-dominance-solvable guessing games, with two or more players facing various (a)symmetric (across players) strategy spaces larger than ours, and with the "winning guess" determined by various sample statistics of players' guesses (e.g., mean, median or maximum) multiplied by various (a)symmetric (across players) target numbers smaller or greater than one. These features jointly determine how one's own guess influences the winning guess, and the number of rounds of iterated elimination of dominated guesses necessary to identify one's iteratively undominated guess(es).[2] Iterated-dominance-solvable guessing games require a simultaneous assessment of players' individual rationality and their beliefs about others' rationality. In dominance-solvable guessing games, by contrast, equilibrium predictions and players' best responses rely only on individual rationality in the game-theoretic sense of obeying simple dominance. This allows us to focus on the basic limits of cognition as revealed by dominance violations.

Behavior in a two-player dominance-solvable guessing game is studied in Grosskopf and Nagel (2007a and 2007b, hereafter GNa and GNb, respectively, or GN).[3] A fixed monetary prize is won by the player(s) whose guess is closest to two-thirds of the pair's average guess. Guesses can range from 0 to 100 (inclusive), so the strategy space is much larger than in our games. Although guessing 0 is a weakly dominant strategy, 90% of subjects (132 undergraduates with no formal training in game theory) initially make dominated guesses above 0, which is close to random guesses (99%). The strikingly frequent dominance violations appear robust to increasing stakes or implementing more detailed explication of the guessing game,[4] and they only partly vanish with more expertise or on-task experience.[5]

GN further offer a comparison of behavior in the two-player dominance-solvable game with behavior in an otherwise identical eighteen-player iterated-dominance-solvable game, played by another 36 subjects from the same population. In the eighteen-player game, only about 10% of subjects initially violate dominance, far below random guesses (33%) and typical for iterated-

[2] See CGC and Grosskopf and Nagel (2007a) for an overview of iterated-dominance-solvable guessing games. Two-player guessing games are normally dominance-solvable but CGC introduce a new class that are iterated-dominance-solvable.
[3] We occasionally use GN to refer to both GNa and GNb which use the same experimental dataset, with GNa analyzing first-round behavior and GNb behavior over time.
[4] This result is based on correspondence with Brit Grosskopf.
[5] Experts (economic researchers at conferences) do better than students but their dominance compliance is still only 37%. GNb further observe that dominance violations persist even after ten rounds of playing the game in fixed pairs: depending on the extent of feedback provided during the game, one- to three-quarters of student subjects guess above 0 in the tenth round.

dominance-solvable games including guessing games.[6] Surprisingly, initial guesses in the two- and eighteen-player games are similar and even marginally higher in the former game with the unique undominated guess of 0. GNb also observe what appears as a lack of knowledge transfer between the two games: In a treatment where subjects switch after four rounds from the eighteen-player to the two-player game, most of them make a higher guess in the fifth round compared to the fourth round.[7]

In the spirit of Charness and Levin's (2007) exploration of the origins of the winner's curse, we implement minimalist dominance-solvable guessing games to look closer at the potential sources of dominance violations. Compared to GN, our two-player games 2p2n and 2p3n constrain the strategy space to only two and three numbers, respectively. In principle, especially game 2p2n permits mentally or visually listing all contingencies – i.e., all combinations of both players' possible choices and their payoff consequences – so one can "gradually recognize" the dominance of choosing 0 even without being a priori aware of the notion of dominance. As illustrated in Figure 1, listing contingencies may be (cognitively) hardest in our three-player game 3p2n. In a game-theoretic sense, however, game 3p2n is not harder than game 2p3n since the former is strict-dominance-solvable while the latter is weak-dominance-solvable.

Although our guessing games are simple, they are neither trivial nor necessarily easier to represent mentally than guessing games with larger strategy spaces. The latter games may facilitate (or necessitate) subjects translating them into stories that are easier to represent mentally, although the high rate of dominance violation in GN's two-player game suggests otherwise. Given GN's findings, we do not expect everyone to solve our games, especially not the arguably more complex games 2p3n and 3p2n. The variation in cognitive and game-theoretic complexity among our games is meant to aid our understanding of the sources of dominance violation. Particularly, the three games may differ in subjects' reasoning processes and reasoning errors.

---

[6] About 90% of subjects guess below ⅔*100, thus seemingly respecting at least one round of iterated dominance. This is typical for initial responses in iterated-dominance-solvable games (see, e.g., CGC's Table 6).

[7] These (and other) comparisons should of course be viewed in light of the different sets of dominated and undominated strategies across the games. Moreover, since subjects played GN's two-player game repeatedly in fixed pairings and knew about it a priori, they might have viewed their first-round guesses as influential for subsequent game play; hence the first-round guesses might not represent true initial responses free of repeated-game effects and experimentation. Also, most of GN's subjects obtained some degree of outcome and payoff feedback which is uncommon in studies of initial responses, though GN document that the distributions of first-round guesses do not differ across their feedback treatments.

Since our focus is on cognition rather than learning, we collect only initial choices in a between-subjects design. Each of our subjects makes a single choice for one of the games depicted in Figure 1, which suppresses any form of learning (including introspective one), repeated-game effects, and experimentation.[8] Nevertheless, the single choice results from a relatively lengthy deliberation process undertaken by subjects when reporting their reasoning. While having only one choice per subject could still undermine the reliability of across-game comparisons of choice behavior, our primary focus is rather on the relationship between subjects' reasoning processes, choices, beliefs, and cognitive and personality characteristics, as detailed below.

3. Reasoning classes, decision-making errors and stated beliefs

In an answer protocol appended to the experiment's instructions, we prompted subjects to report their complete reasoning leading them to their choice, and then to state their choice and beliefs about the choice(s) of the other player(s) in their pair or triplet (see Appendix 1).[9] Subjects were told to report their reasoning in as much detail as possible in order to get paid.

Two examiners from outside the research team – CERGE-EI third-year Ph.D. students with advanced training in game theory – independently classified subjects' reasoning processes based on inspecting copies of the answer protocols, without observing the stated choices and beliefs which we deleted from the protocols. This was to ensure that the examiners focus on classifying subjects' reasoning processes rather than inferring the classification from the stated choices and beliefs, with the ultimate aim of detecting any differences between reasoning processes and choices.[10]

We gave the examiners classification instructions (see Appendix 2 for details) asking them to assign each subject's reasoning process into one of the following three reasoning classes:

[8] In our opinion, our games are too similar to each other to warrant their implementation in a within-subjects design, especially given our parallel elicitation of reasoning processes, choices and beliefs. Even with new partners for each game and no feedback, or even with the games embedded within a set of other games, we would risk considerable introspective learning, making it difficult to disentangle learning from cognition. See CGC for a detailed argument for studying truly initial responses.

[9] Subjects in game 3p2n were reminded that they could state different beliefs about the choices of the other two players in their triplet, but none of them actually did so.

[10] The examiners were of course not completely blind with respect to choices and beliefs since subjects often indirectly stated them as part of their reasoning. However, as will become clear below, such indirect statements could form part of various reasoning processes and had to be carefully interpreted by the examiners in the context of a particular reported reasoning process.

Reasoning class A

Wrong reasoning – e.g., due to misrepresenting the strategic nature of the guessing game or making a numerical mistake, or irrelevant belief-based reasoning.

Reasoning class B

Reasoning based on listing contingencies involving own dominant choice of 0, but without explicitly explaining why 0 is the dominant choice.

Reasoning class C

Reasoning explicitly recognizing and explaining why 0 is the dominant choice, with or without listing contingencies.

Class A includes a variety of wrong reasoning processes discussed in detail in Section 6.3. Class A, for example, includes irrelevant belief-based reasoning such as "I believe the other player chooses 1, so I will choose 1 and we will split the prize."[11] By contrast, belief-based explanations of dominance are included in reasoning class C – e.g., "I believe the other player chooses 0 because that's the best for her, so I will choose 0 not to lose the game," or "I expect the other player to choose between 0 and 1 randomly or with some probabilities, but no matter what she chooses, my best choice is 0."

For class B, listing contingencies means listing the combinations of the pair's or triplet's possible choices and their consequences, in any plausible mathematical, verbal or graphical form. However, since it would have been impossible for the examiners to distinguish between intentional and unintentional omission of (irrelevant) contingencies involving own dominated choices, class B requires listing only the contingencies involving own dominant choice of 0:

Game 2p2n: contingencies involving choice pairs (0, 0) and (0, 1)

Game 2p3n: contingencies involving choice pairs (0, 0), (0, 1) and (0, 2)

Game 3p2n: contingencies involving choice triplets (0, 0, 0), (0, 0, 1) and (0, 1, 1)

Class B therefore includes subjects who used the correct (if not most efficient) approach which in principle allowed them to recognize the dominance of choosing 0, but who apparently did not

[11] We argue in Section 6.3 that irrelevant belief-based reasoning might in fact stem from misrepresenting the strategic nature of the guessing games. There is no indication that irrelevant belief-based reasoning could be induced by our belief elicitation procedure or result from fairness considerations.

recognize it. By contrast, class C includes subjects who explicitly recognized and explained the dominance of choosing 0. In addition to the aforementioned belief-based explanations of dominance, class C subjects used reasoning such as "Choosing 0 is an always-winning choice," or "If I choose 1, I can lose, whereas if I choose 0, I always win or at worst tie," or "If I choose 0, I don't need to take the choice(s) of the other player(s) into account."

To the extent that subjects did not always report their reasoning clearly and completely, we cannot rule out classification errors. If uncertain whether a subject falls into class A (class C), the examiners assigned the subject into a "borderline" class A/B (class B/C). If uncertain whether a subject used erroneous belief-based reasoning or rather a belief-based explanation of dominance, the examiners assigned the subject into a "borderline" class A/C. Appendix 2 outlines further steps taken to minimize classification errors.

We repeatedly reminded the examiners that our primary classification goal was to maximize the accuracy of assignment into reasoning classes A and C. This assignment turns out to be robust in that, except for four and three subjects, respectively, the examiners' independent assignments into class A and class C coincide. The robustness is much lower for class B and the borderline classes where the assignments mostly arise from an initial disagreement between the examiners and subsequent re-classification.[12] In the discussion of results below, we therefore mainly concentrate on the robust classes A and C.

Our classification procedure improves upon previously implemented classifications of reasoning processes in iterated-dominance-solvable guessing games (e.g., Bosch-Domenech et al., 2002; CGC) which relied on correctly disentangling individual (ir)rationality from beliefs about others' (ir)rationality.[13] Our classification, by contrast, focuses solely on whether subjects are rational in terms of obeying simple dominance. Reasoning processes in a dominance-solvable guessing game were also collected by GN but were used only to illustrate specific cases of dominance violation. Our advantage over GN lies in our games having constrained strategy spaces. As a result, our subjects mostly report their reasoning in an easily interpretable manner, which reduces the potential scope for classification errors.

[12] See Appendix 2 for details of the re-classification procedure.

[13] Bosch-Domenech et al. classify mostly optionally reported reasoning processes from lab, classroom, and field experiments. As an implementation caveat, the classification is done by the authors themselves. The authors use the classification to conclude that guess distributions visually differ across reasoning classes broadly as predicted by iterated belief types. CGC collect reasoning processes only ex-post through a debriefing questionnaire, and use them to diagnose reasoning errors (of the kind we discuss in Section 6.3) or exotic decision rules not discernible from subjects' guesses alone.

This has important implications for interpreting the relationship between reasoning classes and choices (Section 6.1). In particular, provided that classification errors are minimal, class C subjects should make the dominant choice of 0, unless they slip up during the ultimate decision-making stage of actually stating their (dominant) choice. By inspecting the choice distribution of class C subjects, we can assess the extent of such choice errors. On the other hand, class A subjects most likely make errors during an earlier reasoning stage of the decision-making process, and we explore the nature of such reasoning errors (Section 6.3) and trace them back to subjects' cognitive and personality characteristics (Section 6.2). We also check the extent to which class A subjects make dominated or (accidentally) dominant choices.

Beliefs about others' choices have no strategic role in our dominance-solvable guessing games as they are theoretically irrelevant for own behavior. For that reason, we do not elicit beliefs in an incentive-compatible manner, and our aim is not to assess whether subjects act on their beliefs (see, e.g., Costa-Gomes and Weizsäcker, 2007). We merely interpret subjects' stated beliefs as an interesting indicator of their view of others' rationality, and we report the beliefs conditional on subjects' own dominance compliance as revealed by their reasoning class and choice (Section 6.1). This usefully complements the evidence from iterated-dominance-solvable games where individual rationality and beliefs are necessarily assessed concurrently.[14]

4. Cognitive, personality and demographic characteristics

Our subjects completed several tests of cognitive abilities, scales measuring personality traits, and a demographic questionnaire. Because of no strong priors regarding which individual characteristics might predict behavior in our games, measuring a broader set of potentially relevant characteristics seemed desirable in order to explore and compare their effect.[15] Below we briefly outline the measured cognitive, personality and demographic characteristics, of which working memory (cognitive ability), need for cognition, and premeditation attitude (personality traits) turn out to be important predictors of subjects' behavior. We refer the reader to Rydval (2007) and Ballinger et al. (2007) for further details of the cognitive tests and personality scales.

[14] We acknowledge that the assessment of beliefs is more informative for class C subjects who apparently understood that others' choices are irrelevant for their own best response, but perhaps less informative for class A subjects who rarely understood the strategic nature of the guessing games (see Section 6.3).

[15] Some of the cognitive tests and personality scales were primarily implemented for the purpose of an unrelated follow-up experiment completed by the subjects (see Section 5 for details).

Working memory is viewed by psychologists as the ability to keep relevant information accessible in memory when facing information interference and to allocate attention among competing uses when executing cognitively complex tasks. Working memory tests proxy general cognitive abilities in that they robustly predict general "fluid intelligence" and performance in a broad range of cognitive tasks requiring controlled (as opposed to automated) information processing (e.g., Feldman-Barrett et al., 2004; Kane et al., 2004). Working memory also positively affects economic performance, such as precautionary saving behavior (Ballinger et al., 2007) or forecasting performance (Rydval, 2007). We measure working memory by a computerized version of the "operation span" test (Turner and Engle, 1989) that requires memorizing sequences (of various lengths) of briefly presented letters interrupted by solving simple arithmetic problems. At the end of each sequence, subjects are asked to recall as many letters as possible in the correct position in the sequence, which in turn determines the test score.

Short-term memory reflects information storage capacity as well as information coding and rehearsal skills that make the stored information more memorable (e.g., Engle et al., 1999). We measure short-term memory by a computerized auditory "digit span" test similar to the Wechsler digit span test (e.g., Devetag and Warglien, 2007). Our test requires memorizing pseudo-random sequences (of various lengths) of briefly presented digits and recalling them immediately after hearing each sequence.[16] The test score is based on the number of digits recalled in the correct position in the sequences.

We measure arithmetic ability using an "addition and subtraction" test under time pressure. The test features alternating rows of two-digit additions and subtractions such as "25+49=__" or "96–24=__," and the test score is the number of correct answers. The test belongs to the class of basic arithmetic skill tests provided by the "ETS Kit of Referenced Tests for Cognitive Factors" (Ekstrom et al., 1976), which proxy the ability to perform basic arithmetic operations with speed and accuracy rather than mathematical reasoning or higher mathematical skills.

We measure subjects' personality traits using several item-response personality scales described below. Personality traits could predict guessing game behavior but could also correlate with measured cognitive abilities, so we measure both to disentangle their effect. Each personality scale consists of a collection of statements (worded positively or negatively) for which subjects

[16] What distinguishes the short-term and working memory tests (and cognitive constructs) is an "attention interference" task in the latter tests, such as the simple arithmetic problems in the operation span test.

indicate their agreement or disagreement on a scale from 1 to 4. The personality scales were included in a single item-response survey in a randomized order identical across subjects.

The need for cognition scale measures intrinsic motivation to engage cognitively demanding tasks (e.g., Cacioppo et al., 1996). There is an extensive (inconclusive) literature in economics and psychology on the channels through which intrinsic motivation could interact with financial incentives in stimulating mental or physical effort and performance (e.g., Deci et al., 1999; Gneezy and Rustichini, 2000; McDaniel and Rutström, 2001; Ariely et al., 2005). Not addressing the complex interactions, we measure intrinsic motivation to account for the possibility that subjects are *ex ante* differentially motivated to solve the guessing games or that intrinsic motivation correlates with subjects' measured cognitive abilities.

The premeditation scale captures the propensity to pause and think carefully while carrying out (cognitive) tasks, which might be relevant for forming sound reasoning processes in our games. The sensation-seeking scale is a general proxy for risk-taking attitude which might affect subjects' willingness to experiment with alternative approaches to solving the guessing games. The perseverance scale measure subjects' determination and perseverance in solving lengthy and demanding tasks.[17] The math anxiety scale is a proxy for feelings of tension when manipulating numbers and solving math problems (e.g., Pajares and Urdan, 1996).

We further elicit risk preferences using a hypothetical "multiple price list" procedure (e.g., Holt and Laury, 2002). While risk preferences should not affect subjects' behavior in our guessing games (since choices have known payoff consequences), risk-taking attitude might matter for the reasons hypothesized above for sensation-seeking. Finally, we administer a demographic questionnaire to collect data on subjects' age, gender, field of study, and socioeconomic status such as (family and personal) car ownership.

5. Implementation details

The experiment was conducted at the Bank Austria Portable Experimental Laboratory at CERGE-EI in November 2005 and January 2006, as displayed in Table 1.[18] The subjects were

[17] The premeditation, sensation-seeking and perseverance scales capture various aspects of impulsive behavior (Whiteside and Lynam, 2001). See Ballinger et al. (2007) for further details.

[18] Due to concerns that subjects in successive experimental sessions might share information relevant for performance in the guessing games and some of the cognitive tests, we ensured to the extent possible that successive sessions overlapped or that subjects in non-overlapping sessions were recruited from different

112 full-time students (Czech natives, with a couple of exceptions permitted based on proficiency in Czech) from Prague universities and colleges, namely the University of Economics, Czech Technical University, Charles University, and Anglo-American College, with the majority of subjects from the first two universities.[19] None of the subjects had prior formal training in game theory.

[Table 1 about here]

Each experimental session started with conducting the cognitive tests and personality scales, followed by the guessing game and the demographic questionnaire.[20] The guessing game itself lasted about 20-30 minutes. We read the instructions aloud (see Appendix 1) and then gave subjects virtually unlimited time to re-read the instructions, to ask any questions, and to fill out the answer protocol. We did not explicitly check subjects' understanding of the instructions. While experimentalists often implement prior understanding tests or unpaid practice rounds to ensure that subjects understand the potential consequences of their and others' decisions, doing so in our simple guessing games would almost inevitably induce undesirable experimenter demand effects or suggest strategies to the subjects.[21] As an alternative to an understanding test, GNa implement more elaborate instructions in several sessions of their dominance-solvable guessing game, but find no impact on behavior.[22] In Section 6.3, we report on additional sessions aimed at gauging the nature and extent of our subjects' misunderstanding.

The experiment lasted 1.5-2 hours and subjects earned 150 CZK (≅ PPP$12) for its completion. In addition, the guessing games featured the fixed prize of M=1500CZK (≅ PPP$117) for the

---

universities or university campuses. Judging from the experiment following the guessing game, subjects' behavior suggests little or no degree of social learning (see Rydval, 2007).

[19] Czech Technical University is a relatively non-selective university mostly offering education in various branches of engineering, while the University of Economics is a more selective university mostly offering education in economics, management and accounting. We do not detect any differences in subjects' behavior related to their field of study, though the sample sizes involved in those comparisons are too small to draw any firm conclusions.

[20] After a short break the sessions continued with an individual decision-making experiment unrelated to the guessing game (a time-series forecasting task; see Rydval, 2007).

[21] Understanding tests of course strive hard to avoid such adverse effects – usually by checking solely that subjects understand how their and others' decisions determine payoffs – but even that may have behavioral consequences. For example, in Bosch-Domenech et al.'s (2002) iterated-dominance-solvable guessing game, subjects who a priori observed an example outlining the consequence of guessing a low number violated dominance less frequently than other subjects not observing that example.

[22] GN explain how the average of the pair's guesses is computed and then multiplied by the target number to determine the winning guess. This explication has no effect on the distribution of guesses, though one should note that the change in instructions coincided with an increase in stakes as well as a minor change in the subject population.

winner(s) originating from one pair or triplet selected at random in each session.[23] All parts of the experiment were anonymous (subjects were assigned a unique ID that they kept throughout the session) and earnings were paid out privately in cash after the experiment. The order of cognitive tests and personality scales was the same across sessions, with the former generally preceding the latter. The working memory and short-term memory tests were computerized using E-prime (Schneider et al., 2002) while the remainder of the experiment was administered in a paper-and-pencil format.

6. Results

6.1 Relationship between reasoning classes, choices and beliefs

Table 2 displays the number of subjects in the reasoning classes defined earlier, aggregated across the three guessing games. The first row shows that 66 subjects (59%) used wrong reasoning processes (class A), whereas 30 subjects (27%) reasoned consistently with dominance (class C). The remaining 16 subjects are scattered among class B and the borderline classes. Thus while class B contains only 3 subjects, it can in principle contain up to 13 subjects depending on how one interprets the borderline classes A/B and B/C. Similarly, class A can contain up to 72 (64%) subjects if adding the borderline classes A/B and A/C, and class C can contain up to 40 (36%) subjects if adding the borderline classes A/C and B/C.

[Table 2 about here]

The second and third rows of Table 2 display the frequencies of dominant and dominated choices. The first column shows that 62 subjects (55%) made the dominant choice of 0 while the remaining 50 subjects (45%) violated dominance.[24] This can be compared to a rate of dominance violation of 56% for random guesses, 10% typically reported for iterated-dominance-solvable games, and 90% for GN's dominance-solvable game (see Section 2). In GN's treatment closest

[23] The guessing games were announced as a bonus task. Subjects knew about the existence of a bonus task (and the potential prize) from initial instructions. Subjects also knew they could earn an additional 900CZK (≅ PPP$70) in the experiment following the guessing games.

[24] For ease of exposition, the Choice=1 category includes the two subjects who in fact chose 2 in game 2p3n. Similarly, the Belief=1 category includes the subject who stated a belief of 2. The prevalence of the dominated choice of 1 over the dominated choice of 2 might signal a focal-number effect of unity.

to ours – because of collecting guesses, beliefs and reasoning processes – 78% of subjects violated dominance.[25]

The second and third rows of Table 2 further show that subjects in class A/C or higher all made the dominant choice of 0: Hence they apparently did not commit any errors at the ultimate decision-making stage of stating their dominant choice. On the other hand, the high frequency of class A and class A/B subjects suggests a prevalence of reasoning errors made at an earlier stage of the decision-making process. Note, however, that little over a quarter of class A and class A/B subjects made the *dominant* choice. Thus to the extent that our classification is correct, the observed frequency of dominated choices in fact understates the actual frequency of dominance violations as revealed by dominance-incompatible reasoning.

The remaining rows of Table 2 display subjects' beliefs, first conditional on reasoning classes and then also conditional on choices. The last column indicates that all but three class C subjects believed that the other player(s) would likewise make the dominant choice. The second column shows that out of the 49 class A subjects who made dominated choices, all but three believed that the other player(s) would likewise do so. On the other hand, out of the 17 class A subjects who made the dominant choice, little over half believed that the other player(s) would likewise do so. This further illustrates that seemingly dominance-compatible choices and beliefs can sometimes be based on dominance-incompatible reasoning processes.

Table 3 disaggregates the percentages of reasoning classes and choices for each game. While our primary interest is not in across-game comparisons, we note that the percentage of class C subjects is highest in game 3p2n – even higher than in the simplest game 2p2n – and lowest in game 2p3n, and vice versa for the percentages of class A subjects. Accordingly, the frequency of dominant choices is highest in game 3p2n and lowest in game 2p3n (and is always higher than random dominance compliance).[26] Thus we might have somewhat "less smart" subjects playing game 2p3n – an issue addressed in the next section – or game 2p3n might be generally harder to solve, perhaps due to its weak-dominance-solvable nature or its larger set of dominated strategies (and hence its lower random dominance compliance).

[25] We are grateful to Brit Grosskopf for providing us with the data for this unpublished treatment. The 78% dominance violation rate is based on a sample of 18 student subjects. We again note design differences between our and GN's games, such as the multi-round nature of their experiment and their payoff function rewarding the winner(s) in each fixed pair in every round.

[26] Nevertheless, the three games are very similar in terms of the proportions of dominant and dominated choices made by class A subjects.

[Table 3 about here]

6.2 Cognitive and personality predictors of reasoning classes and choices

Here we assume in a very simple manner that reasoning errors and choice errors have a logistic structure. In particular, Table 4 reports logit estimates of the effect of statistically relevant cognitive and personality characteristics on reasoning classes and choices.[27] The dummies for game 2p2n and game 3p2n capture any remaining differences with respect to game 2p3n. In all estimations, we drop the three class A/C subjects, leaving us with 109 subjects.

[Table 4 about here]

Model 1 reports marginal effects for ordered logit estimation with the reasoning classes as the dependent variable: We conservatively re-assign subjects from the borderline classes A/B and B/C to classes A and B, respectively. The estimates for the game dummies confirm the overall higher likelihood of sounder reasoning processes in games 2p2n and 3p2n compared to game 2p3n. The remaining estimates suggest that higher working memory, need for cognition, and premeditation are associated with a lower likelihood of reasoning inconsistently with dominance (class A) and with a higher likelihood of reasoning consistently with dominance (class C).[28]

Since a Hausman-type specification test for *Model 1* suggests that treating reasoning classes B and C separately is unnecessary, we merge the classes in *Model 2*. The resulting logit estimates reaffirm the results of *Model 1*, namely the positive predictive power of measured working memory, need for cognition, and premeditation for our subjects' ability to reason consistently with dominance. A one-standard-deviation increase in any of the three variables is associated

[27] Other cognitive, personality and demographic characteristics are individually and jointly insignificant at the 10% level. As an exception, arithmetic ability affects both reasoning classes and choices when included instead of or besides working memory, but its impact is generally weaker than that of working memory. Since we lack arithmetic ability scores for the first experimental session, we focus on the effect of working memory in the full sample, noting that working memory is correlated with arithmetic ability at the 10% significance level (Spearman correlation of 0.19) and hence that part of the explanatory power of working memory may reflect the impact of arithmetic ability. One could in principle separate the impact of working memory, arithmetic ability, and short-term memory on behavior (see Rydval, 2007), but this seems undesirable here due to the limited sample of subjects with arithmetic ability and short-term memory scores (additional ten observations are missing for logistical reasons).

[28] In all estimations, the working memory score is the total number of correctly recalled letters *only* in letter sequences recalled entirely correctly. An alternative score, based on the total number of correctly recalled letters, has less predictive power in our estimations. See Conway et al. (2005) for a comparison of the two valid working memory scoring procedures.

with an increase in the likelihood of using a sound reasoning process (here class B or higher) by over 10 percentage points.[29]

Model 3 reports marginal effects for logit estimation with choices as the dependent variable. The negative estimates reflect that higher working memory, need for cognition, and premeditation are associated with a higher likelihood of making the dominant choice of 0. A comparison of *Model 2* and *Model 3* reveals, however, that the predictive power and magnitude of impact of the three cognitive and personality characteristics is much higher in the former model.[30] This further confirms that reasoning classes, rather than choices, represent a useful indicator of our subjects' ability to reason consistently with dominance.

We note once again the marked across-game differences in reasoning processes and choices which prevail even after accounting for the effect of cognitive and personality characteristics.[31] Though not reported in Table 4, there are no differences in the effect of working memory, need for cognition, and premeditation across the games. Therefore, dominance seems harder to understand or apply in the weak-dominance-solvable game 2p3n (with lower random dominance compliance) compared to the strict-dominance-solvable games 2p2n and 3p2n.

6.3 Further exploring origins of dominance-incompatible reasoning

Our classification procedure reveals further insights about the nature of reasoning errors. First, the examiners indicated that misunderstanding experiment's instructions appeared rare: only up to three class A subjects seemingly misunderstood that they played an iterated-dominance-solvable guessing game against everyone else in their session.[32] Second, nearly a quarter of class A subjects apparently failed to incorporate the target number, ½, in their reasoning process, yet this was unlikely due to their misunderstanding of the instructions *per se* (since we stressed the target number when reading the instructions aloud) but rather due to a reasoning or computational error. Third, nearly half of class A subjects reported irrelevant belief-based reasoning such as "I believe the other player chooses 1, so I will choose 1 and we will split the

[29] These effects are independent to the extent that working memory, need for cognition, and premeditation are not correlated in our subject sample at the 10% significance level.
[30] We include need for cognition in *Model 3* for the purpose of a direct comparison with *Model 2*. Although the preferred model of choice behavior does not feature need for cognition, including it does not affect the significance of the other regressors.
[31] Although need for cognition is on average significantly lower for subjects in game 2p3n than in game 3p2n at the 5% level (using a two-sided rank-sum test and *t*-test), Table 4 shows that this cannot explain the lower performance of subjects in game 2p3n.
[32] Here and hereafter, we quantify the maximum extent of specific types of reasoning errors, as indicated by *either* of the examiners in the nine-class classification scheme (see Appendix 2).

prize," or irrelevant focal-number reasoning such as "I like number 1 more than number 0 and hence choose 1." These kinds of irrelevant reasoning might likewise stem from failing to incorporate the target number, i.e., interpreting it as unity. In the 2p2n and 2p3n games, this would imply a game where both players win regardless of their choices. In game 3p2n, this would imply a game where one's own choice has a pivotal influence on the winning choice only if the other two players choose 0 and 1. Thus choosing 0 would no longer be a dominant strategy in any of our games.

This suggests that many class A subjects might have misrepresented the strategic nature of our guessing games and hence played a wrong game. Indeed, in Devetag and Warglien's (2007) two-player game similar to our guessing games, nearly a quarter of subjects misrepresented the relational structure between own and other's preferences.[33] Strategic misrepresentations could be even more widespread in our games since Devetag and Warglien explicitly display the to-be-represented preferences, whereas we rely on subjects inferring the preference (payoff) structure from the verbal instructions.

To explore the nature and extent of strategic misrepresentations, we conducted four additional sessions for the simplest game 2p2n. We changed our experimental design in that subjects first filled out all contingencies, i.e., the four combinations of both players' possible choices and their payoff consequences (see Appendix 1). We also asked subjects to rank the contingencies according to their preferences, had they been able to choose among them (subjects could express indifference between any contingencies by using ranking such as "1, 2, 2, 4" or "1, 3, 3, 3"). Only then were subjects prompted to report their complete reasoning leading them to their choice and to state their choice and beliefs. The additional sessions otherwise resembled the original ones, including the payoff function and the subject population.[34]

[33] Devetag and Warglien categorize two-player games by type of bi-ordered preference structures varying in relational complexity. Their subjects select four out of 16 possible squares simultaneously representing two order relations, one represented by the size and the other by the color of the squares. Our guessing games are order-isomorphic to antitonically projective preference bi-orders found in games of conflict, where players' preference relations are the reverse of one another (though our games feature non-strict payoff relations unlike Devetag and Warglien's games). Harder relational structures found in chicken games and prisoner's dilemma games were misrepresented by 34% and 52% of subjects, respectively (138 undergraduate and MBA students in two related experiments).

[34] Subjects' earnings did not depend on how they filled out and ranked the contingencies, but completing these tasks was a precondition for receiving the participation fee. Subjects were undergraduates from the University of Economics and Czech University of Life Sciences in Prague.

Before answering our key question, we note that out of the 64 additional subjects, 50% reasoned in line with dominance (class C) while the rest did not (class A), and 31% of subjects violated dominance by making dominated choices. Hence compared to our findings in Tables 2 and 3, asking subjects to represent game 2p2n in terms of its contingencies seems to slightly reduce but certainly not eliminate dominance violations.[35] This finding is in the spirit of GNa's observation that implementing more elaborate instructions in their dominance-solvable guessing game does little to improve subjects' understanding of dominance.

As to our key question, all but one class C subject filled out the four contingencies correctly.[36] By contrast, all but two class A subjects were unable to do so, despite always correctly listing all combinations of choices. A third of class A subjects assigned identical payoff, M/2, to both players in all contingencies (and their reasoning reveals a failure to incorporate the target number, ½); another third assigned the prize, M, to the dominated rather than the dominant choice; and the remaining third filled out partly or entirely wrong payoffs – seemingly illogical fractions of M or payoffs not summing to M (for a given contingency).

7. Discussion and conclusion

To understand the nature of dominance violations in dominance-solvable guessing games, we study the relationship among subjects' reasoning processes, choices, beliefs, and cognitive and personality characteristics. Our classification of reasoning processes suggests that only 27-36% of subjects reason in line with dominance; they all make the dominant choice and almost all expect others to do so. On the other hand, 59-64% of subjects reason inconsistently with dominance, of which about three-quarters make dominated choices (and almost all of those expect others to do so) but the remaining quarter perhaps accidentally make the dominant choice (and half of those expect others to do so).

Our additional findings in Section 6.3 reveal that dominance-incompatible reasoning stems primarily from subjects misrepresenting the strategic nature of the games. Specifically, half of

[35] These observations rest on classification done by the authors. Similar to our findings in Tables 2 and 3, our additional class C subjects all make the dominant choice and almost all believe that others would do so; most of the additional class A subjects make dominated choices and expect others to do so, yet nearly a third of class A subjects make the dominant choice and nearly half of those expect others to do so.

[36] Interestingly, a quarter of class C subjects indicated in their ranking (and sometimes also in their reasoning) a preference for splitting the prize with the other player – by ranking highest the contingencies with choice pairs (0, 0) and (1, 1) – or an indifference between winning and splitting the prize. The remaining class C subjects ranked the four contingencies according to their own payoff.

our additional subjects, and likely a similar fraction of our original subjects, were unable to connect their own and others' choices with the payoff consequences, despite the minimalist nature of our games and despite having virtually unlimited time for clarification questions and for making decisions. This kind of bounded rationality, observed in a similar form by Devetag and Warglien (2007), underlies the game-theoretic notion of "sampling equilibria" (e.g., Osborne and Rubinstein, 1998). We note, however, that this kind of bounded rationality may conceivably be less widespread in more naturalistic settings than ours.

In our original subject sample, the likelihood of reasoning errors – most likely misrepresentation errors – is higher for subjects' with lower ability to maintain and allocate attention, as measured by working memory. This is in line with Devetag and Warglien's (2007) finding of a positive link between short-term memory and the ability to represent preference structures similar to our guessing games – though short-term and working memory are quite distinct cognitive constructs (see Section 4) – and also in line with Burnham et al.'s (2007) finding of a positive link between a short test of general cognitive ability and performance in an iterated-dominance-solvable guessing game.[37] We acknowledge that the observed impact of working memory may partly reflect the influence of arithmetic ability (see footnote 27). Also, the effect of working memory might be a combination of a direct effect on behavior and an indirect one activated by requiring subjects to report their reasoning processes.[38]

In our original subject sample, reasoning (misrepresentation) errors are also more likely for subjects with *ex ante* lower intrinsic motivation and premeditation attitude, presumably due to their lower willingness to engage in solving the guessing games or to carefully think through the solution. In our view, this does not contradict CGC's conclusion that deviations from theoretical predictions in their iterated-dominance-solvable guessing games are mainly driven by cognitive errors rather than insufficient motivation. Our findings suggest that insufficiently (intrinsically)

[37] Neither of the two studies account for other potential sources of individual heterogeneity in cognitive abilities and personality traits (as Devetag and Warglien acknowledge), though Burnham et al. control for individual differences in gender, education, and age.

[38] Cognitive scientists, especially proponents of Protocol Analysis, usually take more care than we did to train subjects in verbalizing thought processes in a manner not interfering with solving the task itself (e.g., Ericsson and Simon, 1993; Ericsson, 2002). Describing thought processes – especially aloud, i.e., not in our case – may require additional cognitive resources, divert task-specific cognitive processes and hence generate invalid descriptions of thoughts, particularly in insight tasks requiring creative thinking (e.g., Schooler et al., 1993).

motivated subjects were most likely excluded from CGC's subject sample after failing an understanding test.[39]

Put differently, our findings suggest that had we implemented the listing of contingencies as an understanding test and dismissed subjects failing it – which would have been mostly those with low working memory, intrinsic motivation or premeditation attitude – we would have observed a much lower rate of dominance violation. From this perspective, our findings do not contradict the much lower dominance violation rates observed in Bone et al.'s (2006) extensive-form dominance-solvable game (where presenting the payoff structure visually presumably makes dominance transparent) or in iterated-dominance-solvable games (where subjects usually have to pass a "payoff structure understanding" test). A similar qualification probably also applies to the dauntingly high dominance violation rate reported in GN's dominance-solvable guessing game.

[39] CGC dismiss about 20% of subjects based on failing a detailed understanding test.

References

Ariely, D., Gneezy, U., Loewenstein G., Mazar, N., 2005. Large stakes and big mistakes. FRB of Boston Working Paper No. 05-11.

Ballinger, P., Hudson, E., Karkoviata, L., Wilcox, N., 2007. Saving performance and cognitive abilities. University of Houston manuscript.

Bone, J., Hey, J. D., Suckling, J., 2006. Do people plan? University of York Discussion Paper No. 2006/22.

Bosch-Domenech, A., Montalvo, J. G., Nagel, R., Satorra, A., 2002. One, two, (three), infinity, ...: newspaper and lab beauty-contest games. American Economic Review 92, 1687-1701.

Burnham, T. C., Cesarini, D., Wallace, B., Johannesson, M., Lichtenstein, P., 2007. Billiards and brains – cognitive ability and behavior in a p-beauty contest. SSE/EFI Working Paper Series in Economics and Finance No. 684.

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., Jarvis, W. B. G., 1996. Dispositional differences in cognitive motivation: the life and times of individuals varying in need for cognition. Psychological Bulletin 119, 197-253.

Charness, G., Levin, D., 2007. The origin of the winner's curse: a laboratory study. American Economic Journal: Microeconomics, forthcoming.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., Engle, R. W., 2005. Working memory span tasks: a methodological review and user's guide. Psychonomic Bulletin and Review 12, 769-786.

Costa-Gomes, M. A., Crawford, V. P., 2006. Cognition and behavior in two-person guessing games: an experimental study. American Economic Review 96, 1737-1768.

Costa-Gomes, M. A., Crawford, V. P., Broseta, B., 2001. Cognition and behavior in normal-form games: an experimental study. Econometrica 69, 1193-1235.

Costa-Gomes, M. A., Weizsäcker, G., 2007. Stated beliefs and play in normal-form games. University of York manuscript.

Deci, E., Koestner, R., Ryan, R., 1999. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychological Bulletin 125, 627–668.

Devetag, G., Warglien, M., 2007. Playing the wrong game: an experimental analysis of relational complexity and strategic misrepresentation. Games and Economic Behavior, forthcoming.

Ekstrom, R. B., French, J. W., Harman, H., Derman, D., 1976. Kit of Factor-Referenced Cognitive Tests (revised edition). Princeton, NJ: Educational Testing Service.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., Conway, A. R. A., 1999. Working memory, short-term memory, and general fluid intelligence: a latent variable approach. Journal of Experimental Psychology: General 128, 309–331.

Ericsson, K. A., 2002. Toward a procedure for eliciting verbal expression of nonverbal experience without reactivity: interpreting the verbal overshadowing effect within the theoretical framework for protocol analysis. Applied Cognitive Psychology 16, 981-987.

Ericsson, K. A., Simon, H. A., 1993. Protocol Analysis: Verbal Reports as Data (revised edition). Cambridge, MA: Bradford books/MIT Press.

Feldman-Barrett, L., Tugade, M. M., Engle, R. W., 2004. "Individual differences in working memory capacity and dual-process theories of the mind. Psychological Bulletin 130, 553-573.

Gneezy, U., Rustichini, A., 2000. Pay enough or don't pay at all. Quarterly Journal of Economics 115, 791-811.

Grosskopf, B., Nagel, R., 2007a. The two-person beauty contest. Games and Economic Behavior, forthcoming.

Grosskopf, B., Nagel, R., 2007b. Rational reasoning or adaptive behavior? Evidence from two-person beauty-contest games. Harvard NOM Research Paper No. 01-09 (revised version).

Holt, C. A., Laury, S. K., 2002. Risk aversion and incentive effects. American Economic Review 92, 1644–55.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., Engle, R. W., 2004. The generality of working memory capacity: a latent variable approach to verbal and visuospatial memory span and reasoning. Journal of Experimental Psychology: General 133, 189-217.

McDaniel, T, M., Rutström, E. E., 2001. Decision making costs and problem solving performance. Experimental Economics 4, 145-161.

Osborne, M. J., Rubinstein, A., 1998. Games with proceduraly rational players. American Economic Review 88, 834-847.

Pajares, F., Urdan, T., 1996. An exploratory factor analysis of the mathematics anxiety scale. Measurement and Evaluation in Counseling and Development 29, 35-47.

Rydval, O., 2007. Financial incentives and cognitive abilities: evidence from a forecasting task with varying cognitive load. Jena Economic Research Paper No. 2007-040 and CERGE-EI Discussion Paper No. 173.

Schooler, J. W., Ohlsson, S., Brooks, K., 1993. Thoughts beyond words – when language overshadows insight. Journal of Experimental Psychology: General 122, 166-183.

Schneider, W., Eschman, A., Zuccolotto, A., 2002. E-prime User's Guide. Pittsburgh: Psychology Software Tools.

Simon, H. A., 1978. Models of Thought, vol. 1. New Haven: Yale University Press.

Simon, H. A., 1989. Models of Thought, vol. 2. New Haven: Yale University Press.

Stanovich, K. E., West, R. F., 2000. Individual differences in reasoning: implications for the rationality debate? Behavioral and Brain Sciences 23, 645-665.

Turner, M. L., Engle, R. W., 1989. Is working memory capacity task-dependent? Journal of Memory and Language 28, 127-154.

Whiteside, S. P., Lynam, D. R., 2001. The five factor model and impulsivity: using a structural model of personality to understand impulsivity. Personality and Individual Differences 30, 669-689.

APPENDIX 1: Instructions and answer protocol

[The instructions below were presented to subjects for game 2p2n. Italics denote alterations for games 2p3n and 3p2n. The instructions were in Czech and were preceded by general instructions explaining, among other things, the anonymity of the experiment and the privacy of the paying-out procedure. Explanatory notes in square brackets do not appear in the instructions. The bold face appears in the instructions.]

BONUS TASK! BONUS TASK! BONUS TASK! BONUS TASK!

ID: _____

In this task, you will be randomly matched with one *(two)* other participant*(s)* in this room who will be solving the same task as you will. The task will be explained below.

From now on, the two *(three)* of you will be called a 'group'.

After everyone has finished the task, the winner from one randomly selected group will earn a prize of 1,500 CZK. If the group has more than one winner, the prize of 1,500 CZK will be split evenly between the winners.

The task:

Each member of the group chooses a number: 0 or 1 *(0, 1 or 2)*.

The winner is the group member whose choice is closest to **½ of the average of the numbers chosen by all group members**. [The experimenter read the instructions aloud, stressing the "½" to ensure the target number was not overlooked.]

[In the additional sessions (see Section 6.3), we inserted here instructions asking subjects to fill out the four combinations of both players' possible choices and their payoff consequences, in four consecutive tables of the following format:

Your choice:		Your payoff:	
His/her choice:		His/her payoff:	

Then subjects were asked to rank the tables (i.e., the combinations of choices and resulting payoffs) according to their preferences, had they been able to choose among them.]

Below, please write down the complete reasoning leading you to your choice and then answer the questions at the bottom of the page. **Write while you think! (If you need more space, please turn over and continue.)**

[Here subjects were given much more space to report their reasoning.]

Your choice:
Number (please circle) 0 1 *(2)*

Question: What choice do you expect from the other member*(s)* of your group?
Number (please circle) 0 1 *(2)*

APPENDIX 2: Details of the classification procedure

In the classification instructions, we first presented the examiners with the three guessing games through a condensed version of the experiment's instructions accompanied by Figure 1. We reminded the examiners that they may encounter different reasoning processes across the three guessing games but that it is important to classify them consistently across the games.

The classification instructions further stressed that "[i]t is extremely important for us that you are consistent in your classification, from the very first to the very last subject. It may well happen during the classification that you change your mind about how you classified a previous subject. This is not an error on your part but please do tell us about such cases before proceeding with further classification." The examiners were encouraged to independently contact the second author in case of any questions or ambiguities, preferably before starting (or restarting) their classification.

To minimize the potential scope for subjective classification errors, we initially asked the examiners to independently classify subjects' reasoning processes according to a more detailed nine-class classification scheme. Being based on our evaluation of reasoning processes in previous pilot experiments, the nine narrower reasoning classes corresponded to the various subtle distinctions among potentially reported reasoning processes (see the above discussion in Section 3).

After the nine-class classification scheme, we were able to clarify those distinctions to the examiners (through examples unrelated to their specific nine-class classification results), to explain them how to classify reasoning processes within the three-class classification scheme and what types of classification errors to attend to. Judging from the examiners' feedback and their classification adjustments between the two classification schemes, we were successful in tackling these issues. Being based on classifying reasoning processes using pre-specified, narrowly-defined nine classes of potentially reported reasoning processes, our classification procedure meets the standards of the Protocol Analysis (see, e.g., Ericsson, 2002).

The three-class classification scheme yielded about 20% of classification disagreements between the examiners, half of which they subsequently jointly resolved (only if they deemed appropriate). This final re-classification procedure therefore left us with 10% (11 out of 112) of classification disagreements, which the examiners jointly assigned into the borderline classes in

accordance with the nature of their disagreement. In similar fashion, the examiners also jointly re-examined the remaining subjects in the borderline classes and re-classified them if they deemed appropriate. For all the above cases, we revealed to the examiners the subjects' stated choices and beliefs to which they were a priori blind.

CAPTIONS FOR FIGURES AND TABLES

Figure 1: The guessing games in normal-form representation

Table 1: Order of experimental sessions and number of participants

Table 2: Frequency of subjects sorted by reasoning classes, choices and beliefs

Table 3: Percentages (rounded to integers) of reasoning classes and choices for each game

Table 4: Logistic regressions of reasoning classes (*Model 1* and *Model 2*) and choices (*Model 3*) on cognitive and personality characteristics

Notes: Marginal effects are evaluated at the means of the regressors. Marginal effects in the ordered logit *Model 1* are for classes A, B and C. Working memory, need for cognition, and premeditation are z-standardized using their sample means and sample standard deviations. Standard errors and tests are based on the heteroskedasticity-robust "sandwich" estimator. *, ** and *** indicate significance of estimates at the 10%, 5% and 1% level, respectively. "Joint significance" stands for a chi-square test of joint significance of working memory, need for cognition, and premeditation. "LR" stands for an approximate likelihood ratio test of the null hypothesis that the coefficients are equal across classes. "Hausman" stands for a Hausman-type specification test of the null hypothesis that classes B and C can be merged.