

JENA ECONOMIC RESEARCH PAPERS



2007 – 092

Three Very Simple Games and What It Takes to Solve Them

by

Ondřej Rydval
Andreas Ortmann
Michal Ostatnický

www.jenecon.de

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich-Schiller-University and the Max Planck Institute of Economics, Jena, Germany. For editorial correspondence please contact m.pasche@wiwi.uni-jena.de.

Impressum:

Friedrich-Schiller-University Jena
Carl-Zeiß-Str. 3
D-07743 Jena
www.uni-jena.de

Max-Planck-Institute of Economics
Kahlaische Str. 10
D-07745 Jena
www.econ.mpg.de

© by the author.

Jena Economic Research Papers 2007-092

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

Ondřej Rydval, Andreas Ortmann and Michal Ostatnický[*]

November 9, 2007

Abstract

We study the nature of dominance violations in three minimalist dominance-solvable guessing games, featuring two or three players choosing among two or three strategies. We examine how subjects' reported reasoning translates into their choices and beliefs about others' choices, and how reasoning and choices relate to their measured cognitive and personality characteristics. Only about a third of our subjects reason in accord with dominance; they always make dominant choices and almost always expect others to do so. By contrast, around 60% of subjects describe reasoning processes inconsistent with dominance, yet a quarter of them actually make dominant choices and a fifth of them expect others to do so. Dominance violations seem to arise mainly due to subjects misrepresenting the strategic nature of the guessing games. Reasoning errors are more likely for subjects with lower ability to maintain and allocate attention, as measured by working memory, and for subjects with weaker intrinsic motivation and premeditation attitudes.

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. We thank Vítězslav Babický, Marián Krajč and Jana Richmanová for helping us conduct the experiment; Randall Engle and Richard Heitz of the Attention and Working Memory Lab at the 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.

[*] Rydval – Max Planck Institute of Economics, Jena, Germany, and 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), Prague, Czech republic. Ortmann and Ostatnický – CERGE-EI. Correspondence to A. Ortmann, CERGE-EI, Charles University Prague and Academy of Sciences of the Czech Republic, Politickych veznu 7, Prague 11121, Czech Republic, Tel: +420-224-005-117, Fax: +420-224-227-143, E-mail: andreas.ortmann@cerge-ei.cz, aortmann@yahoo.com.

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 and Crawford, 2006, for an overview; hereafter CGC). Non-equilibrium behavior is typically attributed to players' beliefs about others' irrationality rather than their own irrationality.[1] Especially CGC, through joint analysis of initial responses and information search patterns in iterated-dominance-solvable guessing games, convincingly demonstrate 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. Grosskopf and Nagel (2007a) report a 90% dominance violation rate in a two-player dominance-solvable guessing game. Devetag and Warglien (2007) show that almost 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, Bone et al.'s (2006) observe only a 5% dominance violation rate in a simple extensive-form game against nature, which suggests that people are in principle capable of applying dominance.[2]

We study 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 confounding effects of other-regarding and risk preferences. Dominance-solvable guessing games have the additional appeal of making beliefs about others' rationality theoretically irrelevant for own behavior.

[1] Specifically, non-equilibrium types of beliefs are estimated assuming that individuals respect iterated dominance but believe others do not respect dominance to varying extents. The resulting belief estimates suggest that individuals respect up to three rounds of iterated dominance, relying on others' dominance compliance much less than warranted by equilibrium predictions.

[2] We refer to the dominance violation rate at the second decision node of Bone et al.'s game, where virtually all of the 152 subjects detect that one distribution of four possible payoffs first-order stochastically dominates another distribution. However, the whole game is in fact a two-stage game, and only about a third of the subjects are able to apply the dominance principle at the first (prior) decision node (using, for example, backward induction or the strategy method).

To better understand decision-making errors that our experimental subjects commit, we ask them to describe in detail their reasoning leading them to their choices, and we also elicit their beliefs regarding others' choices. We then examine how subjects' reported reasoning translates into their choices and beliefs. Following the lead from psychology (e.g., Simon, 1978, 1989) and recently experimental economics (e.g., Ballinger et al., 2007; Rydval, 2007), we also study how reasoning and choices relate to subjects' measured cognitive abilities and personality traits.

Only about a third of our subjects reason in accord with dominance; they always make dominant choices and almost always expect others to do so. By contrast, around 60% of subjects describe reasoning processes inconsistent with dominance, yet a quarter of them actually make dominant choices and a fifth of them expect others to do so. Dominance violations seem to arise mainly due to subjects misrepresenting the strategic nature of the guessing games. 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 attitudes.

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.

Previous studies predominantly focus on iterated-dominance-solvable guessing games, with two or more players facing various (a)symmetric guessing spaces much 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 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).[3]

Figure 1: The guessing games in normal-form representation

Game 2p2n: 2 players, 2 numbers

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

Iterated-dominance-solvable guessing games require a simultaneous assessment of both players' individual rationality and their beliefs about others' rationality. In dominance-solvable guessing games, by contrast, players' best responses and equilibrium predictions rely only on individual

[3] 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 due to asymmetric (across players) guessing spaces and target numbers.

rationality in the game-theoretic sense of obeying "simple" dominance. This allows us to focus on 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, 2007b; hereafter GNa and GNb, or GN).[4] 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 guessing 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 guess above 0. The strikingly frequent dominance violations appear robust to increasing stakes or implementing more detailed explication of the guessing game,[5] and they only partly vanish with more expertise or on-task experience.[6]

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.[7] Surprisingly, initial guesses in the two- and eighteen-player games are similar, and are even marginally higher in the two-player dominance-solvable game where one would expect dominance to be easier to apply and hence guesses to be lower.[8, 9]

We implement minimalist dominance-solvable guessing games in order to look closer at the potential sources of dominance violation. Compared to GN, our two-player games 2p2n and 2p3n constrain the strategy space to only two and three numbers, respectively. In principle,

[4] 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.
[5] This result is based on correspondence with Brit Grosskopf.
[6] 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.
[7] About 90% of subjects guess below ⅔*100, hence respecting at least one round of iterated dominance. This result matches typical findings from iterated-dominance-solvable games including guessing games. See, for instance, CGC's Table 6.
[8] GNb further report lack of experience/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 choose a higher number in the fifth round compared to the fourth round.
[9] Certain aspects of GN's design complicate the comparison of initial responses in the two- and eighteen-player games, as well as comparisons with most previously studied iterated-dominance-solvable games. Specifically, since GN's subjects played the two-player game repeatedly in fixed pairings and knew about it from the start, they might have viewed their first-round guesses as influential for subsequent game play; hence first-round guesses might not represent true initial responses free of repeated-game effects and experimentation. Furthermore, most 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.

especially the simplest game 2p2n permits mentally or visually listing a complete set of contingencies (i.e., combinations of both players' possible choices and the resulting outcomes), so one can "gradually recognize" that choosing 0 is a strictly dominant strategy 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.

Our guessing games are very simple but not entirely trivial. Given the frequent dominance violations documented in GN's two-player dominance-solvable guessing game, 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.

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. Hence we completely suppress any form of learning (including introspective one), repeated-game effects and experimentation.[10] While having only one choice per subject likely undermines 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 beliefs

In an answer protocol appended to the experimental instructions, we prompted subjects to describe their complete reasoning leading them to their choice, and to report their choice and beliefs about the choice(s) of the other player(s) in their pair or triplet (see Appendix 1).[11] Subjects were told to complete the answer protocol in as much detail as possible in order to get paid.

[10] Our games are clearly 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, 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.

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

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 reported 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 reported choices and beliefs, with the ultimate aim to detect any differences between reasoning processes and choices.[12]

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

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 the concluding section. 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." 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 the resulting outcomes (for the guessing game that a subject faces), in any

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

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 gradually recognize the dominance of choosing 0, but who apparently did not 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 worse 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 describe their reasoning clearly and completely, we cannot rule our 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 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.[13] In the discussion of results below, we therefore mainly concentrate on the robust classes A and C.

Our classification procedure arguably improves upon previously implemented classifications of reasoning processes in guessing games. Classification in iterated-dominance-solvable guessing

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

games (e.g., Bosch-Domenech et al., 2002; CGC) likely generates a relatively larger extent of classification errors due to the need to disentangle individual (ir)rationality from beliefs about others' (ir)rationality.[14] 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 describe their reasoning in an easily interpretable manner, which reduces the potential scope for classification errors.

This has important implications for interpreting the relationship between reasoning classes, choices and beliefs. In particular, provided that classification errors are minimal, class C subjects should make the dominant choice of 0, unless they slip up at the ultimate decision-making stage of actually reporting their 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 attempt to relate such reasoning errors to subjects' cognitive and personality characteristics. We also check the extent to which class A subjects make dominated or, accidentally, dominant choices.

While beliefs about others' rationality are theoretically irrelevant for own behavior in our dominance-solvable guessing games, subjects' reported beliefs are an interesting indicator of their view of others' rationality. We assess subjects' beliefs about others' dominance compliance conditional on their own dominance compliance as revealed by their reasoning class. As discussed earlier, such separate assessment of individual rationality and beliefs about others' rationality is possible due to our focus on dominance-solvable games, as compared to iterated-dominance-solvable games where individual rationality and beliefs are necessarily assessed concurrently.[15]

[14] 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 choice 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 mainly to interpret behavior of subjects with unclear or borderline types of iterated beliefs.

[15] The conditional analysis of beliefs is informative predominantly for class C subjects who apparently understood that others' choices are irrelevant for their own best response, but less informative for class A subjects, especially those whose beliefs stem from irrelevant belief-based reasoning.

4. Cognitive, personality and demographic characteristics

Our subjects completed several tests of cognitive abilities, personality scales and a demographic questionnaire. Because of no strong priors as to 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.[16] Below we briefly outline the measured cognitive, personality and demographic characteristics, of which working memory, need for cognition and premeditation attitudes 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.

Working memory is perceived 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 time-series forecasting performance (Rydval, 2007). We measure working memory by a computerized version of the "operation span" test (Turner and Engle, 1989) that requires subjects to memorize 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 better memorable (e.g., Engle et al., 1999). We measure short-term memory by a computerized auditory "digit span" test closely resembling the Wechsler digit span test (e.g., Devetag and Warglien, 2007). The test requires subjects to memorize pseudo-random sequences (of various lengths) of briefly presented digits and to recall

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

them immediately after hearing each sequence.[17] The test score is based on the number of digits recalled in the correct position in the sequence.

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 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 in effortful, 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 attitudes 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

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

demanding tasks. [18] The math anxiety scale is a proxy for feelings of tension when manipulating with numbers and solving math problems (e.g., Pajares and Urdan, 1996).

We further elicit risk preferences using a hypothetical random lottery pair design (e.g., Holt and Laury, 2002). While risk preferences should not affect subjects' behavior in the guessing games (since choices always have known payoff consequences), risk-taking attitudes might matter for reasons hypothesized above for sensation-seeking. Finally, we collect data on subjects' age, gender and field of study, and proxies for socioeconomic status such as car ownership.

5. Implementation details

The experiment was conducted at the Bank Austria Portable Experimental Laboratory at CERGE-EI in November 2005 (seven sessions) and January 2006 (one session), as displayed in Table 1.[19] The subjects were 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, the Czech Technical University, the Charles University, and the Anglo-American College, with a majority of subjects recruited from the first two universities.[20] None of the subjects had prior formal training in game theory.

Table 1: Order of experimental sessions and number of participants

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

[18] 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.

[19] 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 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).

[20] The 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.

Each experimental session started with conducting the cognitive tests and personality scales, followed by the guessing game and the demographic questionnaire.[21] 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.

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 winner(s) originating from one pair or triplet selected at random in each session.[22] 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 structure of the guessing games was publicly announced through the experimental instructions (see Appendix 1). We read the instructions aloud 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 experimental instructions. While experimentalists often implement prior understanding tests or unpaid practice rounds to make sure that subjects realize the potential consequences of their own and others' decisions,[23] doing so in our simple guessing games would almost inevitably induce undesirable experimenter demand effects.[24] Furthermore, GNa actually test for the impact of implementing more elaborate experimental instructions, which – given the simplicity of their dominance-solvable guessing game – is akin to implementing an understanding test, but find no impact on behavior.[25] We return to the issue of subjects' misunderstanding in the concluding section.

[21] 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).
[22] The guessing games were announced as a bonus task. Subjects knew about the existence of a bonus task and the potential prize from the initial experimental instructions. Subjects also knew they could earn additional 900CZK (≅ PPP$70) in the experiment following the guessing games.
[23] For example, CGC require subjects to pass a detailed understanding test and dismiss about 20% of them based on failing the test.
[24] Experimenter demand effects seem to show up in Bosch-Domenech et al.'s (2002) iterated-dominance-solvable guessing game. Some of their subjects were shown an example outlining the consequence of guessing a particular low number, which led to a lower frequency of dominance violations compared to other subjects not observing that example.
[25] GN conduct several sessions where they 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.

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

	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

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.[26] This can be contrasted with the 10% rate of dominance violation typically reported for iterated-dominance-solvable games and the 90% rate of dominance violation in GN's dominance-solvable game (see Section 2). In GN's treatment closest two ours – because of its collection of guesses, beliefs and reasoning processes – 78% of subjects violated dominance.[27]

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

[27] 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 implementation differences between our and GN's design, such as the multi-round nature of their experiment and their payoff function rewarding the winner(s) in each fixed pair in every round.

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 decision-making errors at the stage of reporting their rational, dominant choice. On the other hand, the high frequency of class A and class A/B subjects suggests a prevalence of reasoning errors. Note, however, that over a quarter of them – 17 class A and two 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 also do so. On the other hand, out of the 17 class A subjects who made the dominant choice, nine believed that the other player(s) would also do so. This further illustrates the heterogeneity of class A reasoning processes, especially that apparently rational choices and beliefs can sometimes be based on entirely irrational reasoning processes.

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

	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 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. Correspondingly, the percentage of class A subjects is highest in game 2p3n and lowest in game 3p2n, and similarly for the percentages of dominated choices.[28] Therefore, unless we have somewhat "less smart" subjects in game 2p3n – an issue we address in the next

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

section – game 2p3n seems generally harder to solve, perhaps due to its weak-dominance-solvable nature.

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.[29] 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.

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).[30]

Since a Hausmann-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

[29] All other measured cognitive, personality and demographic characteristics are individually and jointly insignificant at the 10% level. As an exception, arithmetic ability significantly 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 for logistical reasons 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 be attributable to 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).

[30] In all estimations, the working memory test score is the total number of correctly recalled letters *only* in letter sequences recalled entirely correctly. An alternative working memory test score, based simply 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.[31]

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 the magnitude of impact) of the three cognitive and personality characteristics is much higher in the former model.[32] This speaks further in favor of the reasoning classes being a useful indicator of our subjects' ability to reason consistently with dominance.

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

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

Notes: Marginal effects are evaluated at the means of the regressors. Marginal effects in the ordered logit *Model 1* are successively 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.

[31] These effects are independent to the extent that working memory, need for cognition and premeditation are not correlated in our sample at the 10% significance level.

[32] 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.

We note once more the marked across-game differences in reasoning processes and choices which prevail even after accounting for the effect of cognitive and personality characteristics.[33] Not reported in Table 4, there are no differences in the effect of working memory, need for cognition and premeditation across the games. Therefore, the dominance principle seems distinctly harder to understand or apply in the weak-dominance-solvable game 2p3n compared to the strict-dominance-solvable games 2p2n and 3p2n.

7. Discussion and conclusion

To shed light on the nature of decision-making errors in simple dominance-solvable guessing games, we study the relationship among subjects' reasoning processes, choices, beliefs about others' choices, and cognitive and personality characteristics. Our classification of reasoning processes suggests that only 27-36% of subjects reasoned in accord with dominance; they always made the dominant choice and almost always believed that others would do so. On the other hand, 59-64% of subjects reasoned inconsistently with dominance, yet a quarter of them actually made dominant choices and a fifth of them expected others to do so.

We observe a variety of wrong reasoning processes, and our classification procedure reveals further insights about the nature of reasoning errors. The examiners for instance indicated that subjects' misunderstanding of the experimental 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.[34] Almost a quarter of class A subjects clearly failed to incorporate the target number, ½, in their reasoning process. However, this was unlikely due to their misunderstanding of the instructions since we stressed the target number when reading the instructions aloud, but rather due to their momentary lapse of attention or computational error during their reasoning process. Up to four class A subjects made other kinds of computational errors.

Almost a half of class A subjects described irrelevant belief-based reasoning such as "I believe the other player chooses 1, so I will choose 1 and we will split the prize,"[35] or irrelevant focal-number reasoning such as "I like number 1 more than number 0 and hence choose 1." These

[33] 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 markedly lower performance of subjects in game 2p3n.
[34] 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).
[35] It is possible that our belief elicitation misled some (class A) subjects to think that others' choices are relevant for their own choice, which might have induced them to use wrong belief-based reasoning.

kinds of irrelevant reasoning *might* have in fact resulted from failing to incorporate the target number, ½ (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 own choice has a pivotal influence on the winning choice only if the other two players choose 0 and 1. Thus 0 would no longer be a dominant choice in any of our games.

Indeed, many of our class A subjects might have misrepresented the strategic nature of our guessing games and hence played a wrong game. Devetag and Warglien (2007) show that almost a quarter of their subjects misrepresent the relational structure between own and the other player's preferences in a two-player game similar to our guessing games.[36] Misrepresentations could be even more widespread in our games since Devetag and Warglien show subjects the to-be-represented preferences, whereas we rely on subjects inferring the preference (payoff) structure from the verbal experimental instructions.

To examine the nature and extent of strategic misrepresentations, we conducted three additional sessions for the simplest game 2p2n. Subject first filled out all contingencies, i.e., the four possible combinations of both players' potential choices and the resulting payoffs (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 number of contingencies by using ranking such as 1, 2, 2, 4 or 1, 3, 3, 3). Only after that were subjects prompted to describe their complete reasoning leading them to their choice, and to report their choice and beliefs. These additional sessions were otherwise identical to the previous ones, including the payoff function and the subject population.[37]

Out of the 47 new subjects, 62% reasoned in accord with dominance (Class C) while the rest did not (class A), and 28% of subjects violated dominance by making dominated choices. Hence compared to our findings in Tables 2 and 3, asking subjects to represent the 2p2n guessing game

[36] The authors categorize two-player games by types of bi-ordered preference (payoff) 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).

[37] 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 first-year undergraduate students from the University of Economics in Prague, similar to Session 6 described earlier.

in terms of its contingencies seems to reduce but certainly not eliminate dominance violation.[38] As to our main question, while all Class C subjects filled out the four contingencies correctly,[39] all but one Class A subject were unable to do so. Nearly a half of them assigned identical payoff, M/2, to both players in all contingencies, and their reasoning reveals that they failed to incorporate the target number, ½, in their reasoning process. The remaining class A subjects either filled out entirely wrong payoffs not proportional to the prize, M, or equally frequently, they erred by assigning the prize to the player making the dominated rather than the dominant choice.

These insights help us interpret our findings in light of the aforementioned literature. At first glance, our findings are more in line with the daunting evidence about individual rationality in GN's two-player dominance-solvable guessing game, rather than the evidence from Bone et al.'s (2006) simple extensive-form game and iterated-dominance-solvable games (e.g., CGC) where dominance violations appear minor. Nevertheless, our new sessions suggest that had we implemented the listing of contingencies as an understanding test and dismissed subjects failing it, we would have most likely observed much lower frequency of dominance violations in our games. A similar qualification also holds for GN's findings.

However, we clearly would not have wished to exclude subjects who made representation errors and hence played a wrong game, nor subjects who made other kinds of reasoning errors. By contrast, including these subjects enables us to observe the differences in their choices and beliefs, as compared to those who constructed a correct representation of the guessing games and reasoned correctly. Moreover, we are able to trace the representation errors and other kinds of reasoning errors back to subjects' cognitive and personality characteristics.

Particularly, the likelihood of making a reasoning error is higher for subjects' with lower ability to maintain and allocate attention, as measured by working memory.[40] As discussed in the

[38] These observations rest on classification done by the authors. Similar to our findings in Tables 2 and 3, our new class C subjects always make the dominant choice and almost always believe that others would do so, whereas a quarter of the new Class A subjects make the dominant choice and the same proportion expect others to do so.

[39] Interestingly, a quarter of class C subjects indicated in their ranking (and some of them also in their reasoning) that they would prefer to split the prize with the other player – by ranking highest the contingencies with choice pairs (0, 0) and (1, 1) – or that they are indifferent between winning and splitting the prize.

[40] This conclusion is also indirectly supported by Devetag and Warglien's (2007) finding of a positive relationship between subjects' short-term memory and their ability to represent the relational structures similar to our guessing games, though we note that short-term and working memory are quite distinct cognitive constructs. Also, as the authors acknowledge, they do not account for other potential sources of heterogeneity besides short-term memory differences.

previous section, we acknowledge that the effect of working memory may be confounded with the effect of arithmetic ability. Also, the effect of working memory may be a combination of a direct effect on behavior and an indirect effect pertaining to the allocation of attention between reporting the reasoning process and solving the guessing game.[41]

Stronger intrinsic motivation (need for cognition) and premeditation attitudes are also associated with a lower likelihood of reasoning errors, which suggests that reasoning errors are linked to subjects' inherent willingness to engage in solving the guessing games and their propensity to carefully think through the solution. The relevance of intrinsic motivation contrasts with 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.

A natural next step is to ask whether working memory, intrinsic motivation and premeditation attitudes affect different stages of subjects' reasoning process. For example, premeditation might help subjects in the initial stage of correctly representing the guessing game, while intrinsic motivation and working memory might rather help prevent attentional or computational errors during later stages of the reasoning process. We observe no relationship between the specific types of reasoning errors described above and working memory, intrinsic motivation or premeditation attitudes. Nor do we observe any marked variation in the types of reasoning errors across our guessing games, which might otherwise shed further light on the observed across-game variation in behavior. Nevertheless, we recognize the potential benefit of incorporating other indicators of cognitive processes – such as patterns of information search, brain activation or eye fixations – to address these subtle issues in more depth.

[41] Cognitive scientists, especially proponents of the 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) may require additional cognitive resources, divert task-specific cognitive processes and hence generate invalid descriptions of thoughts, especially in insight tasks requiring creative thought (e.g., Schooler et al., 1993).

References

Ariely, D., Gneezy, U., Loewenstein G., and N. Mazar (2005), "Large Stakes and Big Mistakes," FRB of Boston Working Paper No. 05-11.

Ballinger, P., Hudson, E., Karkoviata, L., and N. Wilcox (2007), "Saving Performance and Cognitive Abilities," University of Houston manuscript.

Bone, J., Hey, J. D., and J. Suckling (2006), "Do People Plan?" University of York Discussion Paper No. 2006/22.

Bosch-Domenech, A., Montalvo, J. G., Nagel, R., and A. Satorra (2002), "One, Two, (Three), Infinity, ...: Newspaper and Lab Beauty-Contest Games," *American Economic Review*, 92, 1687-1701.

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., and W. B. G. Jarvis (1996), "Dispositional Differences in Cognitive Motivation: The Life and Times of Individuals Varying in Need for Cognition," *Psychological Bulletin*, 119, 197-253.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., and R. W. Engle (2005), "Working Memory Span Tasks: A Methodological Review and User's Guide," *Psychonomic Bulletin and Review*, 12, 769-786.

Costa-Gomes, M., and V. P. Crawford (2006), "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study," *American Economic Review*, 96, 1737-1768.

Deci, E., Koestner, R., and R. Ryan (1999), "A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation," *Psychological Bulletin*, 125, 627–668.

Devetag, G., and M. Warglien (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., and D. Derman (1976), *Kit of Factor-Referenced Cognitive Tests* (revised edition), Princeton, NJ: Educational Testing Service.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., and A. R. A. Conway (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., and H. A. Simon (1993), *Protocol Analysis: Verbal Reports as Data* (revised edition), Cambridge, MA: Bradford books/MIT Press.

Feldman-Barrett, L., Tugade, M. M., and R. W. Engle (2004), "Individual Differences in Working Memory Capacity and Dual-Process Theories of the Mind," *Psychological Bulletin*, 130, 553-573.

Gneezy, U., and A. Rustichini (2000), "Pay Enough or Don't Pay at All," *Quarterly Journal of Economics*, 115, 791-811.

Grosskopf, B., and R. Nagel (2007a), "The Two-Person Beauty Contest," *Games and Economic Behavior*, forthcoming.

Grosskopf, B., and R. Nagel (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., and S. K. Laury (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., and R. W. Engle (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., and E. E. Rutström (2001), "Decision Making Costs and Problem Solving Performance," *Experimental Economics*, 4, 145-161.

Pajares, F., and T. Urdan (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 2007-040 and CERGE-EI Discussion Paper 173, 01/2007.

Schooler, J. W., Ohlsson, S., and K. Brooks (1993) "Thoughts beyond Words – When Language Overshadows Insight," *Journal of Experimental Psychology*: *General*, 122, 166-183.

Schneider, W., Eschman, A., and A. Zuccolotto (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.

Turner, M. L., and R. W. Engle (1989), "Is Working Memory Capacity Task-Dependent?" *Journal of Memory and Language*, 28, 127-154.

Whiteside, S. P. and D. R. Lynam (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: Experimental instructions and answer protocol

[The instructions below were presented 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 with 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.]

[Here we inserted the extra instructions for the three additional sessions (see Section 7), asking subjects to fill out four tables with the four possible combinations of both players' potential choices and the resulting payoffs. The four tables had the following format:

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

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

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 writing space.]

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 experimental 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 discussed in the text.

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, one-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 accord 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' reported choices and beliefs to which they were a priori blind.