Non-strategic players are the rule rather than the exception

José de Sousa, Guillaume Hollard and Antoine Terracol*

October 23, 2013

Abstract

Independent studies on experimental games provide converging, and puzzling, evidence: a noticeable fraction of players behave in a non-strategic way. It is hard to believe that so many players act in way that is hard to reconcile with any theory of strategic behavior. This study is designed to understand why. We test commonly suggested explanations such as insufficient stakes, lack of attention or misconceptions about the game. None appear to be entirely convincing. Using reaction-time data, we show that non-strategic subjects spend time thinking about the games and do pay attention to changes. Moreover, players classified as non-strategic in a first set of games continue to act non-strategically in subsequent games. Our results suggest that the existence of non-strategic players in one-shot games is a robust feature of human cognition. Bearing in mind that our subjects are international chess players, we wonder why their strategic chess-playing ability does not transfer to laboratory games. Theoretical and economic consequences are discussed.

*José de Sousa: Université Paris-Sud, ADIS and Sciences-Po, Paris, LIEPP, jose.de-sousa@u-psud.fr. Guillaume Hollard: Paris School of Economics and CNRS, guillaume.hollard@univ-paris1.fr. Antoine Terracol: Paris School of Economics and Centre d'Économie de la Sorbonne, Université Paris 1 Panthéon-Sorbonne, terracol@univ-paris1.fr. We thank G. Aldashev, J. Andreoni, G. Attanasi, A. Caplin, G. Coricelli, M.P. Dargnies, G. Frechette, T. Gajdos, H. Maafi, G. Mayraz, M. Niederle, T. Offerman, J. Sobel, L. Santos Pinto and participants at the 2012 ASFEE conference, the "Decision and Experiments" group at Paris school of Economics, the MBB on-going research seminar, the NYU CESS seminar, the SBDM 2012 in Paris, the Schumpeter Seminar in Berlin, the EEA 2013 conference, the PEJ 2013 conference and the economic seminars in Rennes, Copenhagen, Paris Sud and Ecole Polytechnique for comments. A. Fischman and E. Kaddouch provided excellent research assistance. N. Bonzou and A. Clauzel from the French Chess Federation provided decisive help. Part of this research was conducted while G. Hollard was visiting Stanford.

1 Introduction

In many experimental games some players behave in a rather non-strategic way. More precisely, they exhibit behaviors that are difficult to account for using reasonable beliefs (e.g. the use of dominated strategies). As a consequence, non-strategic players typically lose money as they do not respond to the monetary incentives provided in lab experiments. Recent studies aimed at better eliciting individual beliefs offer additional and puzzling evidence: the fraction of non-strategic players is found to be much higher than initially thought, ranging from 20 to 80%.[1]. This high fraction sometimes came to a surprise to the authors themselves. They were thus rather skeptical about their own results. It is indeed hard to admit that such a large fraction of subjects can fail so dramatically in experimental games. The present study is designed to understand why.

Non-strategic players can be viewed in two rather opposite ways. On the one hand, folk explanations suggest that non-strategic players simply do not pay attention to the instructions of the games. According to this view, non-strategic players are considered as noise or errors. On the other hand, if the impossibility to act strategically in one shot games is observed in a series of independent studies, we are maybe facing a strong regularity: some players lack strategic ability. For instance, they may be unable to transfer to a new strategic context their ability to anticipate others behavior. To discriminate between these two explanations, we design an experimental protocol, with chess players as subjects, that consists of three phases. The data from each of these three phases in turn gradually paint a portrait of non-strategic players by eliminating some folks explanations and providing new evidence.

Overall, the present paper presents five main contributions. First, we propose a method of identifying non-strategic players which controls for beliefs. Using phase 1 data, a series of 10 beauty-contest games,[2] we find a considerable proportion of non-strategic

[1]See Agranov, Caplin, and Tergiman (2013), Agranov, Potamites, Schotter, and Tergiman (2012), Brown, Camerer, and Lovallo (2012), Burchardi and Penczynski (2011), Costa-Gomes and Weizsäcker (2008), Fragiadakis, Ivanov, Knoepfle, and Niederle (2012), Ivanov, Levin, and Niederle (2010) and Ostling, Wang, Chou, and Camerer (2011) Fehr and Huck (2013).

[2]The beauty-contest or guessing game is fairly simple, as described by Nagel (1995): a large number of players have to state simultaneously a number in the interval [0, 100]. The winner is the person whose chosen number is closest to the mean of all of the numbers chosen multiplied by a common knowledge

players, of almost one half of the sample. These players appear not to take into account relevant information relative to their strategic environment. Although, they are not the main focus of the present study, it is important to note that the remaining subjects act in accordance with game theory and seem to anticipate the existence of a large fraction of non-strategic players.

Second, we test whether our strategic vs non-strategic classification has any out-of-sample predictive power in explaining the data from phases 2 and 3, which include a different game. These data from phases 2 and 3 confirm that players who were classified as non-strategic in phase 1 continue to act non-strategically in subsequent games. Moreover, even when all possible precautions have been taken to ensure that they understand the rules and instructions of the game, the players continue to act randomly.

Third, using reaction-time data, we show that non-strategic subjects spend time thinking about the games. They *do* pay attention to relevant changes in the environment but *fail* to take these changes into account. Even when the stakes are raised, non-strategic players are unable to process information in a relevant manner.

Furthermore, since our subjects are chess players recruited during an international tournament,[3] we can safely rule out the possibility that non-strategic subjects are simply unable to act strategically in general, or suffer from deficits in cognitive ability. Indeed, the Elo ranking – which measures the ability to play chess – does not have any significant influence on the likelihood of being classified as non-strategic. Last, using phase 3 data we can see whether possible misconceptions about the reality of the cash payment play an important role in our field experiment (as subjects are participating in an experiment for the first time).

Our results overall suggest that the existence of non-strategic players in one-shot games is a robust feature. Rather than disregarding non-strategic players as noise, we

positive parameter m. For $0 \leq m < 1$, there is a unique Nash equilibrium in which all of the players announce a value of zero.

[3]International Chess players were used in this experiment as we thought that they should exhibit some minimal ability to play games, i.e. they should be able to figure out that they are facing an opponent. We are fully aware that chess players may not be very different from usual lab subjects, even if this point is subject to controversy (see the discussion in Levitt, List, and Sadoff (2011) and Palacios-Huerta and Volij (2009), as well as the evidence from a beauty-contest game in Bühren and Frank (2010)).

believe that they should be considered as one of the main empirical regularities found in situations in which economic agents are confronted with new situations. We thus examine the theoretical consequences of our findings in conclusion.

The remainder of the paper is organized as follows. Section 2 presents the experimental design, and Section 3 discusses the phase 1 results, which allow us to draw a first sketch of non-strategic players. Section 4 then exploits phase 2 data to test the robustness of our preliminary conclusions and render our portrait of non-strategic players more detailed, and Section 5 appeals to the phase 3 data to complete our portrait. Last, Section 6 concludes.

2 Experimental design

We recruited 270 chess-players during a major international tournament held in Paris in 2010. Subjects were approached while they were at the tournament (but not playing). They were then allocated to an adjacent room that serves as an experimental lab. The experiment was computerized. All players read the instructions on the screen; these were also read aloud by the experimenter. Subjects were allowed to ask questions.

Our experiment consisted of three phases:

Phase 1. Subjects were asked to play a series of 10 beauty-contest games. They had to choose a number as close as possible to $m\times$mean (where mean indicates the mean of the answer of all players). The parameter m took on two values: $m = 2/3$ and $m = 4/3$.

Each game was played against five types of opponents, labeled as A, B, C, D and Random. The letters indicates the Elo-Ranking of the opponent who were thus explicitly identify as chess-players.[4] "Random" indicates that the subject is facing a random device that will select a strategy using a uniform probability distribution over the strategy space. Subjects played 10 different games: one game against each type of opponent (A, B, C, D or Random) for each value of $m \in \{2/3, 4/3\}$. The order of the ten games was randomized. Subjects received no feedback during the 10 games.

[4]The letters correspond to the following ranking: A=Elo $\geq$ 2150, B=2150 $>$ Elo $\geq$ 1800, C=1800 $>$ Elo $\geq$ 1500, D=Elo $<$ 1500.

All treatments were identical except that half of the subjects played the beauty contest against *two* opponents of the same level (i.e. A, B, C, D or Random), while the other half played against *one* opponent only. This difference matters as the two-player version of the game has a dominant strategy, while the three-player version does not. In addition, as the payment rule is the same (10 points for each of the ten games, to be shared among the winners), there is a difference in expected earnings: those who play the three-player version of the game earn 33.33 points on average, while those playing the two-player version of the game earn 50 points on average.

Phase 2: After playing their ten beauty contests, subjects start a new game, the 11-20 game (described below). They played this game only once, against another chess player whose level was not specified. We added questions to ensure that subjects had understood the rules before starting the game.

After completing the eleven games (the ten beauty contest plus the 11-20 game), the screen displayed the numbers chosen by players in the 10 beauty-contest games and in the 11-20 game. Subjects were given the opportunity to observe the consequences of their actions. Each action was associated with a number of points and these points were in turn converted into Euros according to a previously announced exchange rate of .2€ per point. Subjects then proceeded in turn to another room where they individually and anonymously received their payments in cash.

Phase 3: *After* receiving their cash payment, subjects were offered the chance to take part in an additional beauty-contest game (with $m = 2/3$) involving all of the participants in the experiment. Subjects were informed that the name of the winners would be publicly announced at the end of the chess tournament. The tournament lasted for 10 days, so players had to wait up to 10 days before receiving their payment were they to win this last game. The two best players (i.e. the two closest to the winning number) both received a cash payment of 150€. These results were publicly announced immediately after the official announcement of the results of the chess tournament. As our subjects were not the usual lab subjects, we were worried that they might not believe that they would really be paid. We thus proposed this additional game *after* they had received their first cash payment, so as to make our promise of a further cash payment credible. Note that at this stage players had also received some feedback regarding the results of their actions in the first two phases.

2.1 Theoretical predictions

We use two games in this paper. The first, the beauty-contest game, is well-known to experimentalists and its theoretical and empirical properties have been well-described. We thus restate the main results. The second game used was recently introduced by Arad and Rubinstein (2012): we will thus present this game in more detail.

2.1.1 The beauty-contest game

The beauty-contest game has been widely used in game theory to capture the notion of step reasoning (see Bühren, Frank, and Nagel (2012) for a historical account). Each player i in this game chooses a number x_i between 0 and 100. The goal is to choose the x_i that is the closest to the target of $m * (\sum_{i=1}^{n} x_i)/n$, where m can take on different values and n designates the number of players. The player whose x_i is the closest to the target wins a fixed prize, while the other players receive nothing.

For $m < 1$, the unique equilibrium in the beauty-contest game is where all players choose to play 0.[5] We will also consider a version of the game with $m > 1$. In this case, the focal equilibrium is that where all players choose 100.[6]

One interesting feature of the beauty contest is that there is a (weakly) dominant strategy with only two players: this strategy is to play 0 when $m < 1$ and 100 when $m > 1$. However, with three or more players there is no longer any dominant strategy.

In the popular case where $m = 2/3$, the mean value chosen in the literature is around 35, which is far removed from the equilibrium prediction. Almost no subjects are found

[5]Note that all players playing 1 may also be an equilibrium if the strategy space is restricted to integers. It is also important to specify that in the case of a tie, either players share the prize or the prize is randomly allocated to one player (in our case, we broke potential ties randomly). If all players receive the entire prize in the case of a tie, additional equilibria may exist

[6]When there are three or more players, there also exists an unstable equilibrium in which all players play 0. This equilibrium no longer exists when there are only two players. See López (2001) for more details on the equilibrium set for integer games.

to play the equilibrium strategy in one-shot games. A variety of different subject pools have played this game, including chess players, with results remaining fairly stable across groups regarding the small numbers who play the equilibrium.

2.1.2 The 11-20 game

The 11-20 money-request game was recently introduced by Arad and Rubinstein (2012) and presented to their subjects as follows:

> "You and another student are playing a game in which each player requests an amount of money. The amount must be an integer between 11 and 20 shekels. Each player will receive the amount he requests. A player will receive an additional 20 shekels if he asks for exactly one shekel less than the other player. What amount of money would you request?"

This game is different from similar games, like the traveler's dilemma, in that to win a prize you have to play *exactly* one step lower than the other player. Given the structure of the game, there is no Nash equilibrium in pure strategy. Assuming both players to be expected gain maximizers, there is a unique symmetric mixed-strategy Nash equilibrium.[7] The symmetric equilibrium distribution puts a zero probability on strategies 11 to 14, probability 1/4 on strategies 15 and 16, and probabilities 4/20; 3/20; 2/20 and 1/20 on strategies 17, 18, 19 and 20, respectively. This equilibrium distribution is not at all obvious to identify, and depends on the assumptions made regarding players' utility functions.

To the best of our knowledge, this game has only been played with students. Arad and Rubinstein found that even students who are trained in game theory do not play as theory predicts. However, their student results do provide a benchmark for the behavior of subjects who are expected to be amongst the most strategic.

[7] There are four other asymmetric mixed strategy equilibria.

3 A portrait of non-strategic players

Using data from phase 1, we first present a criterion to classify players as strategic or not and then discuss its pros and cons. A natural question is the degree to which non-strategic players look like random players. Random players, also known as level-0 players in reference to level-k models,[8] are defined as players who simply pick up a strategy at random out of the strategy space. We thus explore whether non-strategic players behave like random players. Last, we test whether the most common assumptions put forward to explain the existence of non-strategic players can explain our results. We end up rejecting these. Extrapolating from the collected evidence, we propose a portrait of non-strategic players.

3.1 Looking for non-strategic players

Strategic play involves two things: forming correct beliefs and optimizing with respect to beliefs. Our design includes a control on beliefs when players are confronted with a random device. Most deviations from best-response should then be attributed to some kind of mistake in optimizing. However, we are agnostic about the nature of the mistakes. This is why the most relevant criterion in our context is to compare strategies used in the "random" condition when m changes in the beauty contest games (phase 1). In particular, playing *lower* when m=4/3 than when m=2/3, against the same random device, indicates a behavior than can hardly be accounted for. We then extend this criterion to games played against humans. It seems indeed reasonable to assume that strategic players **will *not* play lower when m=4/3 than when m=2/3** while they are facing the same opponents.

Since some aspects of this approach may be debatable, we will discuss more in depth these issues in a specific section. We first present details about our criterion. We thus count the number of times each individual played lower with m=4/3 than with m=2/3,

[8]These models were first presented in Stahl and Wilson (1994, 1995) and Nagel (1995) and have given rise to numerous publications including Camerer, Ho, and Chong (2004), Crawford and Iriberri (2007) and Crawford, Costa-Gomes, and Iriberri (2013)

when facing the same type of opponent.[9] As there are five pairs of games against the same opponents (level A, B, C and D, and Random), there are five observations per player, which we can use to calculate a "Pairwise Rationalizable Actions Index" (PRAI). The players are thus distributed into six categories: an index value of 0 indicates that the subjects systematically play a lower number when m=4/3 than when m=2/3; a value of 5 indicates that no such violation of consistency occurred. Table 1 shows the distribution of players according to this index.

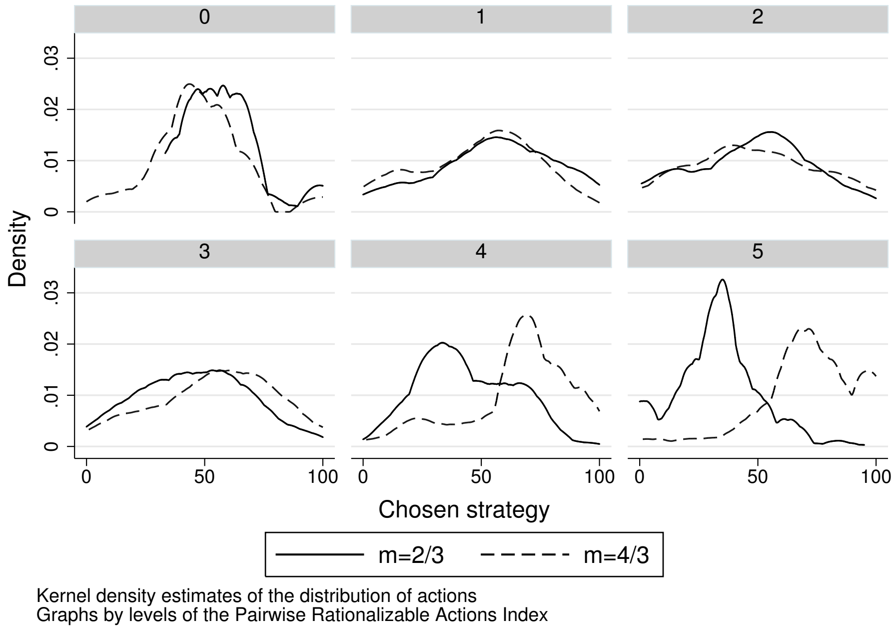
Table 1: Classification of players according to PRAI

Index value	N	Frequency
0	15	5.6
1	21	7.8
2	26	9.6
3	61	22.6
4	46	17.0
5	101	37.4
Total	270	100.0

Our empirical strategy is first to identify a subgroup of players who fail in some respect to act strategically and then carry out various tests to investigate the behavior of this particular sub-population in more detail. In order to streamline our analysis, we regroup some categories. As can be seen from Figure 1, the behavior of players across games is comparable for those with a PRAI value of 0 to 3 versus those with values of 4 or 5, for ease of exposition, the first group will be referred to as "non-strategic" and the second as "strategic". This is a slight abuse of language since so far we can only claim that non-strategic subjects *sometimes* use strategies that can not be rationalized, i.e. in some cases their behavior cannot be though of as being a best-response to their subjective beliefs. But, as the next sections will explain, there are good reasons to think that these subjects are non-strategic in a broader sense.

[9]For the sake of completeness, it is possible that some players hold very extreme beliefs about their opponents in the three-player version of the game which allow them to rationalize playing lower in the 4/3 than in the 2/3 version of the game. This is very unlikely to occur but however remains theoretically possible.

Figure 1: Distribution of actions by PRAI level



3.2 PRAI index: interest and limitations

Our criterion has some pros and cons. It has the advantage to avoid appealing to any complex belief-elicitation mechanism, which would not be convenient given that our experiment was designed to last no longer than 20 minutes. However, some aspects are debatable. We review here four important points.

Alternative criterion Had we identified non-strategic players as those using dominated strategies, we would have missed out a considerable percentage of non-strategic players (56 out of our 123 non strategic players did not play dominated strategies). Using information from a series of 10 games allows us to collect more information on each player and led to a more accurate classification.

Alternative way of splitting our sample into two As any classification is debatable, it is worth asking whether alternative cut-off points of the PRAI index lead to similar results. The behavior of categories 0 to 3 appears to be similar, **while different** from categories 4 and 5, suggesting that our classification is meaningful. **However**, alternative ways of grouping categories may be considered and corresponding results are available upon request.

Endogeneity The reader might worry that we use a criterion that is based on observed choices in phase 1, and that our results for phase 1 data merely come from the way we split our sample, and not from any underlying difference between players. It could well be the case that all players simply randomize their actions, but that some end-up as high PRAI merely by chance. However, we provide simulations, presented in Appendix A, that show that the distribution of our sample according to the PRAI index would have been different had all players simply picked strategies at random.

Extension to other games Our criterion is in part specific to our design and we do not deny an arbitrary aspect. However, comparing the behavior *across* games provides relevant information regarding individual behavior. The usual classification strategy is to classify players according to their behavior in a single game. Here, we suggest that we can check whether players react in the expected direction when a parameter changes. This admittedly does not apply to any pair of games, but nor does the dominated strategy criterion.

3.3 Non-strategic players as random players

The beauty-contest game has been played many times, with various kinds of subject pools. The mean value for the 2/3 version of the game is often found to lie between 35 and 38, with a standard deviation ranging from 20 to 25. Our subjects play slightly higher than the usual lab subjects (students) at close to 42. We obtain similar results in the 4/3 version of the game, with chess-players playing slightly lower. Our results are not particularly high: Camerer, Ho, and Chong (2004) discuss experiments in which more extreme values are sometimes observed (e.g. they report a mean value of 54 in the 2/3 version of the game) and Agranov, Caplin, and Tergiman (2013) find similar figures, especially when players have a limited time to think about the game (30 seconds).

We split our sample into two subgroups based on the index described above. We first compare the behavior of non-strategic and strategic players as the key parameter of the game changes from 2/3 to 4/3. One group, our non-strategic players, does not react as m varies. Non-strategic players, on average, play something close to the salient value of 50, whatever the value of m (48.98 when $m = 2/3$ vs. 51.25 when $m = 4/3$). In sharp contrast, strategic players react in the expected direction as m changes. They play on average 36.12 when $m = 2/3$ and 70.09 when $m = 4/3$.[10] Their average behavior is roughly in line with that of players who best-respond to opponents who select their strategy using a random distribution with a mean of 50.[11] In that sense they resemble the description of level-1 players.

Figure 2: Strategies chosen by low and high PRAI players

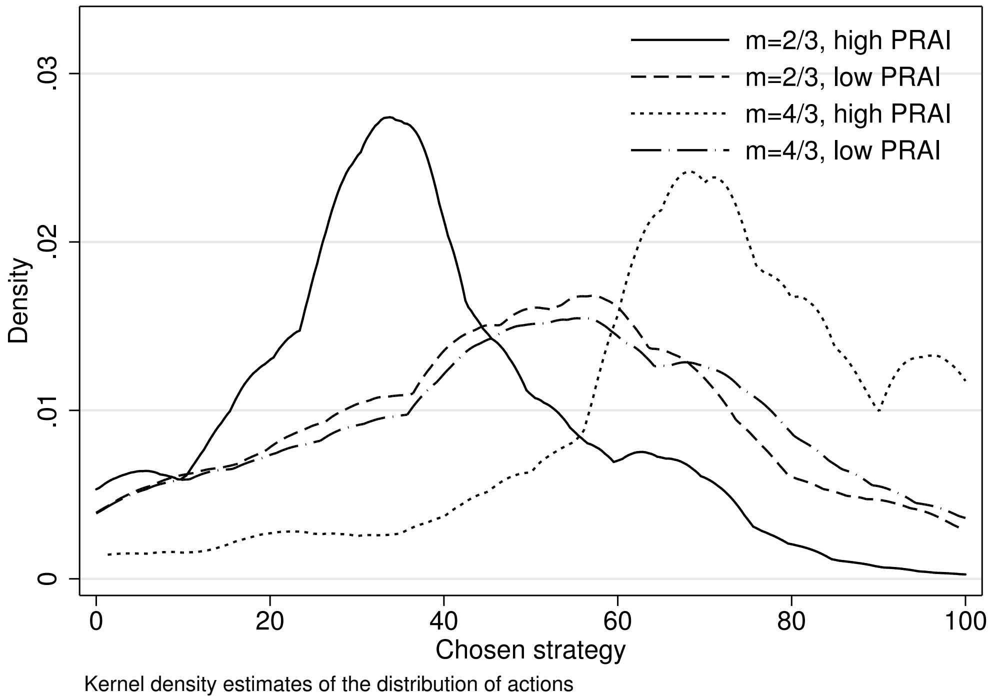


Figure 2 shows these differences between the two groups. That the four distributions appear in that order stems from the way the two groups were constructed (high PRAI individuals play higher at $m = 4/3$ than at $m = 2/3$ by definition). This figure is however

[10]Descriptive statistics on the choices of strategic and non-strategic players can be found in Tables 8 and 9 in Appendix C. Fixed-effects regressions of choices on a dummy for $m = 4/3$, controlling for opponent type and period, yield an estimated coefficient of 2.19 (p-value=0.046) for non-strategic players and of 33.91 (p-value10^{-3}) for strategic players. The difference in the explanatory power is also striking, with a within R-squared of 0.018 for non-strategic players and 0.528 for strategic players.

[11]More precisely, by roughly we mean that strategic players behave as if they were playing a best-response against level-0 players, but they make an often-observed mistake: they fail to take into account the fact that their own choice is included in the calculation of the mean. A value of 33 is a best response to players playing 50 only in games which involve a large number of players.

informative regarding the difference when m varies for the non-strategic players. Indeed these players seem not to react at all to a change of m. We thus suspect that they play in a rather random manner. We thus next examine whether non-strategic players behave like level-0 players. Level-0 players are assumed to pick a strategy randomly from the strategy space without any further consideration of the rules of the game or their opponents' strategy. It is possible that all the players currently classified as non-strategic have adopted a deterministic strategy that they use in all ten games. As shown in Table 2, we can rule out this possibility. The two panels in this table show the correlation coefficients between the five choices for each value of m. The choices are numbered in the order in which they are played. The coefficients for strategic subjects (Table 2, top panel) range from 0.649 to 0.791, while those for non-strategic players (Table 2, bottom panel) are much lower and range from 0.154 to 0.372. Moreover, for non-strategic players, the correlation between choices falls as the time between choices rises, which is not the case for strategic players. Overall, non-strategic players seem to pick strategies (almost) randomly, while strategic players appear to be much more consistent across games.

Table 2:
Correlations of choices over time: strategic vs non-strategic players

	Choice 1	Choice 2	Choice 3	Choice 4	Choice 5
Strategic players					
Choice 1	1.000				
Choice 2	0.696	1.000			
Choice 3	0.705	0.790	1.000		
Choice 4	0.648	0.721	0.745	1.000	
Choice 5	0.666	0.740	0.744	0.782	1.000
Non-strategic players					
	Choice 1	Choice 2	Choice 3	Choice 4	Choice 5
Choice 1	1.000				
Choice 2	0.358	1.000			
Choice 3	0.214	0.280	1.000		
Choice 4	0.217	0.153	0.283	1.000	
Choice 5	0.277	0.199	0.241	0.371	1.000

Interpretation: The correlation coefficient between the values chosen the first and second time the subjects played games with same value of m is 0.696.

We may well wonder whether acting in a non-strategic way affects Phase 1 earnings. As can be seen from Table 3, earnings increase in an almost monotonic fashion with the

Table 3: Earnings in Euros by PRAI level

Index value	Earnings Mean	Earnings Std. Dev.	p-value
0	3.6	1.9	
1	4.7	1.7	0.109
2	5.1	2.0	0.451
3	5.5	1.8	0.287
4	5.4	2.2	0.773
5	7.2	2.7	0.000

The p-values reflect the t-test of a difference in earnings relative to the previous index level

value of the PRAI index. Players with an index value of 0 make only half as much as do those with an index value of 5 (3.6 vs 7.2). As expected, acting non-strategically entails considerable financial losses. Non-strategic players do spend time thinking (see below) but are literally leaving money on the table by not acting strategically.

As a result, we can rule out the possibility that many subjects are using heuristics which we do not understand but which are nonetheless effective.[12] Were some strange but effective strategies to have been used by more than a small fraction of players, we would have seen less significant differences in earnings across groups.

3.4 Testing common assumptions for the existence of non-strategic players

The reason why some players perform so poorly in experiments is not currently well-understood. In what follows we test the most common assumptions that have been advanced in the literature. Broadly speaking the most common of these is that non-strategic players do not exert effort or pay attention. This could be because they think that the stakes are too low or because they are simply unable to perform the task required. In what follows, we consider these assumptions in the light of the specific features of our protocol.

[12] For example, recent evidence suggests that a fair proportion of subjects in the guessing game - a game similar to the beauty contest - do use some kind of rule that is not described in standard game theory but which however makes sense, as in Fragiadakis, Ivanov, Knoepfle, and Niederle (2012)

3.4.1 Do non-strategic players simply not pay attention?

It is difficult to know whether lab subjects are paying attention or exerting effort. Our data do however offer some insights into this question. We recorded the reaction time in the beauty-contest games: these offer a guide to the cognitive effort exerted by subjects. For instance, subjects who do not want to excert any effort or pay attention may play much faster than subjects who spent time thinking about the game and came up with relevant strategies. At the other extreme, if subject are not focusing on the game but on something else, they again should exhibit reaction times that are different from the ones of players who exhibit the most strategic behavior.

First, non-strategic players spend slightly *more* time thinking about the problem than do the other players: the former spend on average 28.87 seconds on each decision, as compared to 25.38 seconds for the latter. The p-value of the t-test on this difference at the individual level is 0.059.[13] This difference in reaction times also holds for the 11-20 game in phase 2, as level-0 players also spend more time thinking (196.83 seconds vs 175.1; p-value=0.029).

The second, and most surprising, results refer to reaction times as the parameter m changes. The tendency across games is for subjects to play faster and faster. However, at some point they are confronted with a change in the parameter m. Recall that the order of the ten games is randomized. Some players will for instance play three games in a row in which $m = 2/3$ and then play the fourth with $m = 4/3$. When subjects are confronted with this kind of change, they need to adapt to a new environment and so increase their reaction time compared to the previous game. Intuition suggests that non-strategic players, were they not paying attention, would not be affected by these changes, as their strategy does not vary with m. However, we actually find that non-strategic players do also react to changes in the value of m. A fixed-effects regression of the reaction time on a dummy for the first change and its interaction with our strategic/non-strategic classification, controlling for period, shows that being faced with the first change in m increases reaction time by 5.39 seconds (p-value=0.066) for strategic players, and

[13]The recorded time for the first decision includes the time taken to read the instructions. We thus do not have a meaningful reaction measure for the first decision.

that this increase is not significantly higher for non-strategic players (the interaction term attracts a coefficient of 1.16 with a p-value of 0.802).[14] This evidence strongly suggests that non-strategic players are aware that something has changed, but that they fail to take this change into account.

3.4.2 Are the stakes too low for subjects to provide any effort?

One possible confound here is that the stakes are too low, so that the expected gains are too slight to make it worth exerting any effort. Our experiment lasted about 20 minutes and subjects were already on site (mostly chatting or hanging around). They received on average about \$16 (11€) for these 20 minutes. Given that subjects did not pay any transport costs, the earnings correspond to an hourly wage of \$48 (33€) which is not different from usual earnings in the lab and perhaps even somewhat higher. We may however continue to wonder about the importance of stakes.

Our design also allows us to test whether expected earnings have any effect. We vary the number of players but keep the winner's reward constant. As such, players in the two-player version of the game have a higher expected payoff (50 points, i.e. 10€) than those in the three-player version (33.3 points, i.e. 6.66€). We thus calculate our index separately for these two populations. We found no significant differences between the distribution of players according to our index, with the proportion of non-strategic players being the same: we classify 45.52 percent of subjects as non-strategic in the two-player version of the game, versus 45.59 percent in the three-player version. The non-strategic percentage is thus not sensitive to a 50 percent rise in expected payoffs. Were stakes to explain the existence of non-strategic players, their fraction would have risen significantly.

We can therefore rule out the possibility that non-strategic players deliberately ignored the incentives as they considered them to be too low.

[14] Also note that the period in which the first change takes place is the same for both player types (2.53 vs 2.62, p-value=0.43).

3.4.3 Are non-strategic players simply unable to think strategically?

The reason behind our recruitment of chess players during an international tournament was exactly to rule out the possibility that subjects could not think strategically. Whatever ability is required to play chess, players by definition have to think about what their opponent will do. We are therefore sure that our subjects, including those who we classify as non-strategic, are actually able to think strategically. Perhaps surprisingly, the Elo ranking plays only a limited role, if any: the average Elo ranking is very similar across our index. Non-strategic players have a mean Elo ranking of 1768 (with a standard deviation of 30), with corresponding figures for strategic players of 1814 (27). This difference is not significant (p=0.25).

3.5 Lessons from phase 1: A portrait of non-strategic players

Evidence from phase 1 data allows us to paint a portrait of non-strategic players. First, we provide evidence that "non-strategic players" behave very much in line with other definition of non-strategic players. For instance, non-strategic players approximately behave as level-0 players since every thing goes on as if they randomly pick a strategy using a symmetric distribution centered at 50 (cf. Table 2). The fraction of non-strategic players found is also very much in line with the estimated fraction of level-0 players found from the estimation of Camerer, Ho, and Chong (2004)'s cognitive-hierarchy model. Corresponding estimations are presented in Appendix B.

Common explanations (lack of attention, limited cognitive ability, insufficient stakes) do not appear sufficient to explain our data. Non-strategic players seem to do their best to play the games (they spend time thinking about them and do understand when the game changes). At this point, our best guess is that non-strategic players do perceive the necessary information, but fail to process it in an appropriate way. Even though our subjects are chess players, everything goes on as if a large proportion of them is unable to retrieve or adapt rules they used in the past to deal with strategic interactions.

Table 4: Number of attempts required to successfully answer 4 questions, by PRAI level

Level	Mean no. of attempts	Median no. of attempts	SD
high PRAI	4.74	4	1.09
Low PRAI	5.67	6	1.40

4 Phase 2 data: robustness checks

The previous section, which covered the first phase, offered a number of insights into the behavior of non-strategic players. Phase 2 of our experiment consists of a single play of the "11-20" game described above. This phase 2 data will allow us to (1) have more control over the instruction phase and (2) test whether our classification from phase 1 is robust across games (i.e. can we make out-of-sample predictions?). We address these two issues in separate subsections.

4.1 Are the instructions producing non-strategic players?

Poor instructions may lead a significant fraction of subjects to misunderstand the rules of the game. In phase 1 we used standard instructions: explanations were both displayed on the screen and read aloud by the experimenter, and subjects had the opportunity to ask questions. In that respect, our instructions are standard. However, we may want to have more control over whether subjects pay attention to the instructions.

The 11-20 game was selected because the instructions are short (only a few lines on the screen) and simple to understand. We further added four comprehension questions, to which subjects had to provide correct answers in order to be allowed to proceed to the game. In particular, subjects were asked to state the payoff of each player in different situations. This allows us to safely rule out the possibility that players did not understand the rules of the game, and thus to limit any misconceptions.

Table 4 shows that non-strategic players require significantly more attempts to provide these four correct answers. Non-strategic players made something like two mistakes (their median number of attempts is six) while most strategic players answered the four questions correctly (with a median value of four).[15] This result reinforces our interpretation of a structural difference between strategic and non-strategic players. Non-strategic players have more trouble learning the rules of a simple game than do other subjects. As such, non-strategic players seem to be slow learners.

One intriguing fact is that the existence of non-strategic players probably has something to do with the instruction phase. Costa-Gomes and Crawford (2006) used very long instructions and found almost no level-0 players in their sample. On the contrary, Georganas, Healy, and Weber (2013) used shorter instructions and found a substantial fraction of level-0 players. Our experiment uses minimal instructions in phase 1, but much more detailed instructions for phase 2, including some quiz questions to detect subjects who did not understand the rules. However, despite our efforts to train subjects, non-strategic players still act randomly (as will be explained below). This raises subtle questions about the way in which subjects should be instructed. Existing evidence (e.g. Chou, McConnell, Nagel, and Plott (2009)) suggests that instructions do matter but their impact is still not perfectly understood.

4.2 Is our classification robust across games?

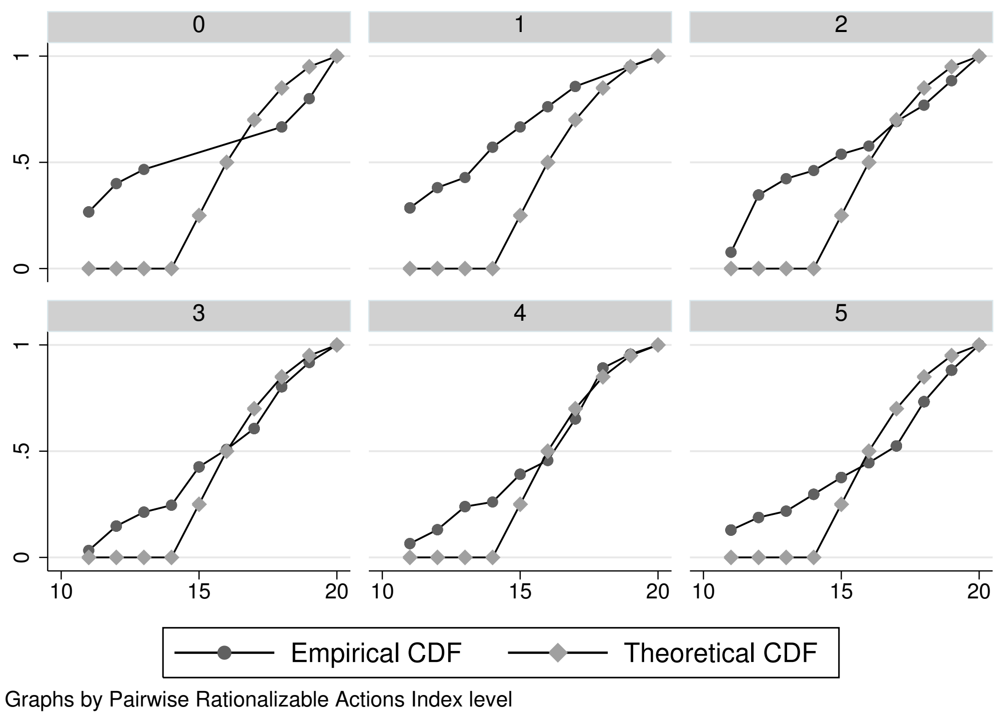
The stability of players' strategic levels across games is a fairly open question. For instance, should we expect a level-2 player in one game to behave as such in a subsequent game? Burnham, Cesarini, Johannesson, Lichtenstein, and Wallace (2009), found that players with a low IQ are much more likely to play dominated strategies in the beauty-contest game and to be classified as level-0, suggesting that being a level-0 player might be a relatively stable individual characteristic across games. However, recent evidence in Brañas Garza, García-Muñoz, and González (2012) has shown that the cognitive-reflexion test is associated with an identifiable pattern in the beauty-contest game, while another, the Raven test, is not. Other authors, such as Georganas, Healy, and Weber (2013), have found only little consistency across games. It is rather unclear whether this absence of any empirical regularity reflects the true absence of stability across games. It could also

[15] Differences in mean and median are both significant at the 1% level.

be the case that the way in which players are assigned to a level is debatable, or that the definition of the levels is problematic. In particular, different definitions of what is level-0 lead to different definitions of higher types, and it is not always easy to assess the stability of levels across games.[16]

To test whether our classification from the beauty-contest games has predictive power for behavior in the 11-20 game, we consider the behavior of each group separately. Figure 3 depicts the empirical cumulative distribution function (CDF) for each of our six PRAI levels, as well as the equilibrium CDF. It is clear from the figure that, although players at all levels fail to play the equilibrium strategy, they are closer to doing so as their index level increases. To test this more formally, we run a probit regression where the dependent variable is 1 if the subject chose an action which is not part of a mixing strategy in the Nash equilibrium. The results are displayed in Table 5, and indeed show that the probability of playing such an action falls with our index.

Figure 3: Cumulative density functions in the 11-20 game



[16]We can here refer to an ongoing project by Bhui and Camerer (2011) which proves that the simple correlation across two games played by the same player may be too demanding a test for stability across games. They suggest rather the use of something like Cronbach's α.

Table 5: Probit regression of out-of-equilibrium actions

Variable	Coefficient	(Std. Err.)
PRAI level	-0.125*	(0.052)
Intercept	-0.021	(0.195)
N	270	
Log-likelihood	-167.473	
$\chi^2_{(1)}$	5.671	

Significance levels : † = 10 %, $*$ = 5 %, $**$ = 1 %

It is also the case that the cumulative distribution of the strategy chosen by low-level players is not distinct from the uniform distribution, while it is so for the high values (i.e. 4 or 5). More precisely, running a Chi-squared test against the discrete uniform distribution over $\{11, 12, \ldots, 20\}$, we find a p-value of 0.22 for low-level players and 0.0003 for high-level players.

To the best of our knowledge, there are not many existing results confirming that strategic sophistication can be used to make out-of-sample predictions, especially when the games are different. Two remarks are however in order. First, we do not make sharp predictions, but only predict that non-strategic players will still behave "randomly". Second, our predictions are based on the outcomes of ten games. In that sense, we use a lot of more information compared to predictions based on only one game, as in Georganas, Healy, and Weber (2013).

4.3 Lessons from phase 2: A clearer portrait of non-strategic players

Adding evidence from phase 2 allows us to provide more detail. Non-strategic players seem to have difficulty in learning in an abstract environment. Even carefully-designed instructions are not enough to generate more strategic behavior from the players who were classified as non-strategic in phase 1.

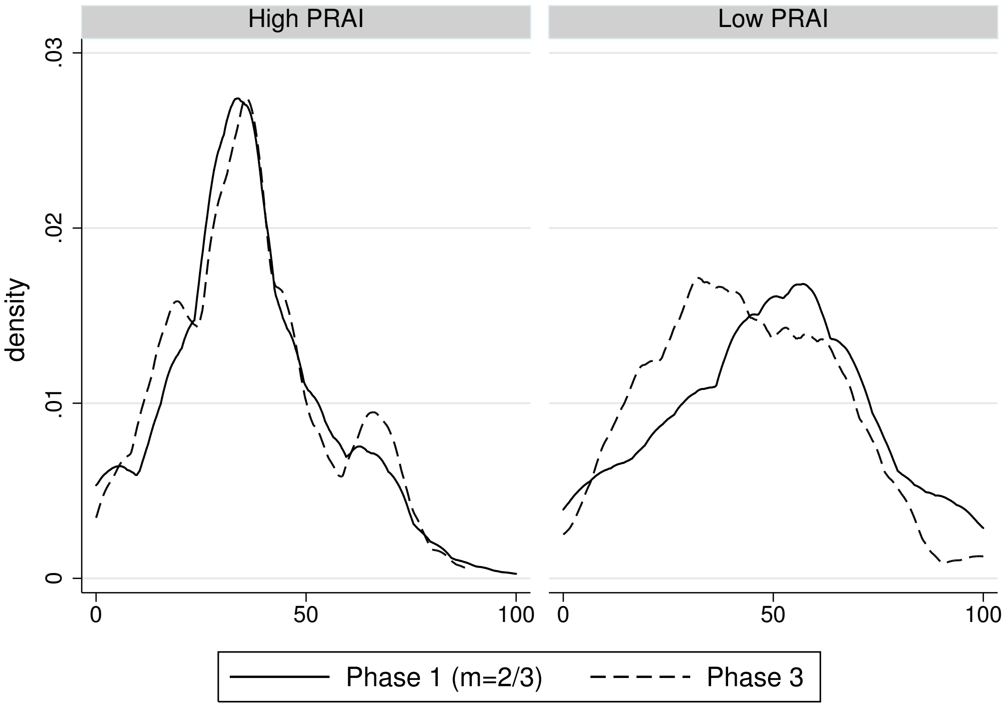
5 Phase 3: increasing stakes and feedback

Phase 3 was intended to make sure that subjects believe that the cash payment will be implemented. One last game was thus proposed after they had received their payment. To reinforce the idea that this was a serious proposition, we also informed them that results would be publicly announced at the same time as the results of the chess tournament. Furthermore, we raised the stakes to about \$200 (150€).

As good practice in experimental economics recommends that subjects can check that their payment corresponds to the announced rules, they had the opportunity at the end of stage 2 to observe the consequences of their actions.

The overall effect was limited (Figure 4) as subjects play 39.4 on average here, compared to an average of 42 in phase 1 when $m = 2/3$. The graphs show that strategic players used very similar strategies to those that they used in phase 1. Non-strategic subjects nevertheless improved a bit. So even if it is not possible to attribute this slight improvement to one particular feature of the game, we can claim that none of the introduced changes had a noticeable impact. We here compare games that are very similar, but which however have some notable differences (e.g. the number of players is not the same and the stakes are different). However, the two groups are still significantly different (a Student test yields a p-value of 0.0139, and a Kolmogorov-Smirnov test a p-value of 0.004). As a conclusion, we can claim that there is no simple trick that would allow us to greatly enhance the degree of sophistication of our subjects.

Figure 4: Actions in Phase 1 ($m = 2/3$) and Phase 3



6 Conclusion

Behavioral models have been proposed in the literature to get a better description of experimental data. However, level-k and quantal response models rest on *ex post* calibration at the subjects' pool level. Attempts were made to get a more precise understanding of individual behavior, leading to various classification of players. The present paper focuses on specific category of players, the less strategic ones. In line with a growing body of independent studies, we find that a sizable proportion of our players, who are chess players, fails to satisfy some minimal strategic requirements. This is in line with Levitt, List, and Sadoff (2011). We have here provided an additional step by testing a number of explanations for the considerable proportion of non-strategic players identified in the literature. The commonly suggested assumptions regarding non-strategic players are lack of attention, misconceptions, insufficient stakes and limited cognitive ability. We show that none of these is a convincing explanation: non-strategic players spend time thinking about the games and do pay attention to changes. When confronted with a simple game (i.e., easy to learn and understand), they continue to act randomly. Even strong chess players may end-up being classified as non-strategic.

So, what prevents chess players to act strategically in experimental games? Non-strategic players understand that something should be done in response to changes in their environment, but have a hard time finding out what that is. So their behavior ends up being radically different from the behavior of players who do something like best-responding, based on their subjective beliefs. Our best explanation regarding this behavior is that these players did not find a way to transfer their strategic ability to experimental games. For instance, case-based reasoning would suggest that facing a new situation, subjects will adapt strategies they used in the past. But even chess players may fail to do so. One can object that most pathologies are observed in one shot games. Repeating the game will certainly reduce these. However, it worth noting that models such as reinforcement learning -one of the most empirically relevant model in game theory- do not assume that players act strategically. It might well be the case that a great part of subjects exhibits a behavior that can be rationalized while they in fact act as if they were facing a problem of decision under uncertainty.

References

Agranov, M., A. Caplin, and C. Tergiman (2013): “Naive Play and the Process of Choice in Guessing Game,” mimeo.

Agranov, M., E. Potamites, A. Schotter, and C. Tergiman (2012): “Beliefs and Endogenous Cognitive Levels: an Experimental Study,” *Games and Economic Behavior*, 75(2), 449–463.

Arad, A., and A. Rubinstein (2012): “The 11-20 Money Request Game: A Level-k Reasoning Study,” *American Economic Review*, 102(7), 3561–73.

Bhui, R., and C. F. Camerer (2011): “Measuring Intrapersonal Stability of Strategic Sophistication in Cognitive Hierarchy Modeling,” mimeo.

Brañas Garza, P., T. García-Muñoz, and R. H. González (2012): “Cognitive effort in the Beauty Contest Game,” *Journal of Economic Behavior & Organization*, 83(2), 254–260.

Brown, A. L., C. F. Camerer, and D. Lovallo (2012): “To Review or Not to Review? Limited Strategic Thinking at the Movie Box Office,” *American Economic Journal: Microeconomics*, 4(2), 1–26.

Bühren, C., and B. Frank (2010): “Chess players’ performance beyond 64 squares: A case study on the limitations of cognitive abilities transfer,” MAGKS Papers on

Economics 201019, Philipps-Universitt Marburg, Faculty of Business Administration and Economics, Department of Economics (Volkswirtschaftliche Abteilung).

Bühren, C., B. Frank, and R. Nagel (2012): "A Historical Note on the Beauty Contest," MAGKS Papers on Economics 201211, Philipps-Universitt Marburg, Faculty of Business Administration and Economics, Department of Economics (Volkswirtschaftliche Abteilung).

Burchardi, K. B., and S. P. Penczynski (2011): "Out of Your Mind: Eliciting Individual Reasoning in one shot Games," mimeo.

Burnham, T. C., D. Cesarini, M. Johannesson, P. Lichtenstein, and B. Wallace (2009): "Higher cognitive ability is associated with lower entries in a p-beauty contest," *Journal of Economic Behavior & Organization*, 72(1), 171–175.

Camerer, C., T.-H. Ho, and J.-K. Chong (2004): "A Cognitive Hierarchy Model of Games," *Quarterly Journal of Economics*, 119(3), 861–898.

Chou, E., M. McConnell, R. Nagel, and C. Plott (2009): "The control of game form recognition in experiments: understanding dominant strategy failures in a simple two person guessing game," *Experimental Economics*, 12(2), 159–179.

Costa-Gomes, M. A., and V. P. Crawford (2006): "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study," *American Economic Review*, 96(5), 1737–1768.

Costa-Gomes, M. A., and G. Weizsäcker (2008): "Stated Beliefs and Play in Normal-Form Games," *Review of Economic Studies*, 75(3), 729–762.

Crawford, V. P., M. A. Costa-Gomes, and N. Iriberri (2013): "Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications," *Journal of Economic Literature*, 51(1), 5–62.

Crawford, V. P., and N. Iriberri (2007): "Level-k Auctions: Can a Nonequilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions?," *Econometrica*, 75(6), 1721–1770.

Fehr, D., and S. Huck (2013): "Who knows It is a game? On rule understanding, strategic awareness and cognitive ability," Discussion Papers, Research Unit: Economics of Change SP II 2013-306, Social Science Research Center Berlin (WZB).

Fragiadakis, D., A. Ivanov, D. Knoepfle, and M. Niederle (2012): "Identifying Predictable Players," mimeo.

Georganas, S., P. J. Healy, and R. A. Weber (2013): "On The Persistence of Strategic Sophistication," mimeo.

Ivanov, A., D. Levin, and M. Niederle (2010): "Can Relaxation of Beliefs Rationalize the Winner's Curse?: An Experimental Study," *Econometrica*, 78(4), 1435–1452.

Levitt, S. D., J. A. List, and S. E. Sadoff (2011): "Checkmate: Exploring Backward Induction among Chess Players," *American Economic Review*, 101(2), 975–90.

López, R. (2001): "On p-beauty contest integer games," Economics Working Papers 608, Department of Economics and Business, Universitat Pompeu Fabra.

Nagel, R. (1995): "Unraveling in Guessing Games: An Experimental Study," *American Economic Review*, 85(5), 1313–26.

Ostling, R., J. T. Wang, E. Y. Chou, and C. F. Camerer (2011): "Testing Game Theory in the Field: Swedish LUPI Lottery Games," *American Economic Journal: Microeconomics*, 3(3), 1–33.

Palacios-Huerta, I., and O. Volij (2009): "Field Centipedes," *American Economic Review*, 99(4), 1619–35.

Stahl, D. O., and P. W. Wilson (1994): "Experimental evidence on players' models of other players," *Journal of Economic Behavior & Organization*, 25(3), 309–327.

——— (1995): "On Players' Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, 10(1), 218–254.

Appendices

A Simulating our Pairwise Rationalizable Actions Index (PRAI) for a homogeneous population

In this Appendix, we run simulations to assess whether the observed distribution of PRAI could have arisen by chance from a homogeneous population. We assume that the population is homogeneous and only composed of random players randomly drawing their actions from the joint empirical distribution of actions. More specifically, each run of the simulation creates 270 individuals. For each individual we draw 5 pair of actions. Each pair is drawn from the empirical joint distribution of pairs of actions against each type of opponent (i.e. A, B, C, D or Random). We thus end up with 5 pairs of actions for each simulated individual. We then calculate the proportion of simulated players falling into each level of our index. We use 9999 runs of the simulation and report the mean value, 1st and 99th percentile of the simulated proportions in Table 6.

Table 6: Simulated versus actual proportions

	Simulated proportions			
Index	1st percentile	Mean	99th percentile	Observed proportion
0	0	0.2	1.1	5.6
1	0.7	2.8	5.6	7.8
2	8.5	13.2	18.1	9.6
3	24.4	30.9	37.8	22.6
4	29.3	36.0	42.6	17.0
5	11.9	16.8	22.2	37.4
Total	-	100.0	-	100.0

As shown in Table 6, the proportions differ substantially if we consider the mean value of each of our 9999 draws. Even if we concentrate on the most unlikely scenarios, the proportion of level 5 exceeds 22.2 percent in under 1 percent of cases, which is far from the observed proportion of 37.4 percent. In our view, this is in line with the fact that our most strategic players are not just random players who happened to draw good strategies by chance. Note that our simulations use the *joint empirical distribution* of pairs of actions; i.e. the scenario most likely to give rise to a simulated distribution very similar to ours.

B Estimations of the (Poisson) Cognitive-Hierarchy model

The cognitive hierarchy model is commonly used to estimate the distribution of players across levels. In what follows, we estimate the Camerer, Ho, and Chong (2004) Cognitive-Hierarchy model where players are distributed among k levels of a cognitive ladder according to a Poisson distribution with parameter τ. Level-0 players play randomly on the strategy space, and level-k, $k \geq 1$ assume that they are the only player at this level, and that the other players are distributed on the levels below according to a normalized Poisson distribution. Players then best-reply to their beliefs over the distribution of players. If $\overline{X}_i$ is player i's belief about the mean action taken by the $N-1$ other players in the game, then his best reply can be shown to be $\frac{N-1}{N-m}m\overline{X}_i$.

We follow Camerer *et al.* (2004) and estimate τ via a method of moments estimator as shown in Table 7. The first two columns list the estimated Poisson parameter and the associated standard error. We estimate the model on various sets of games. The third column shows the implied proportion of level-0 players.

The estimation using the complete sample (i.e. pooling all of the data) yields the prediction that 72 percent of players are at level-0. We obtain similar results when we apply the model to restricted samples of games. The cognitive-hierarchy model thus predicts a very considerable fraction of level-0 players, with at least two-thirds of players being classified as such.

Table 7: Estimation of the cognitive hierarchy model

Sample	$\hat{\tau}$	Std. Err	Percent level-0
All	.32	.03	72
$N = 2$	.29	.03	74
$N = 3$	.39	.04	68
$N = 2$; $m = 2/3$	.40	.08	67
$N = 2$; $m = 4/3$	.27	.04	76
$N = 3$; $m = 2/3$	.42	.09	66
$N = 3$; $m = 4/3$	.38	.06	68

Notes: Standard errors obtained by block-bootstrapping the estimates. Subjects had to choose a number as close as possible to m times the mean of the answer of all players. N refers to the number of opponents faced by each subject.

C Descriptive statistics in the ten beauty-contest games

Table 8: Means and standard deviations for non-strategic players

		$m = 2/3$		$m = 4/3$	
Other players	Obs	Mean	Std. Dev.	Mean	Std. Dev.
A	123	51.6	24.3	50.1	26.6
B	123	51.6	25.1	50.7	25.5
C	123	48.6	24.7	54.7	24.3
D	123	48.3	24.8	47.4	25.8
Random	123	44.9	23.2	53.4	25.3
Overall	615	48.98	24.47	51.25	25.54

Table 9: Means and standard deviations for strategic players

		$m = 2/3$		$m = 4/3$	
Other players	Obs	Mean	Std. Dev.	Mean	Std. Dev.
A	147	34.8	20.5	71.0	23.7
B	147	35.0	16.9	72.3	21.3
C	147	35.9	18.6	71.4	21.3
D	147	38.1	20.9	68.7	21.7
Random	147	36.8	18.1	67.0	21.5
Overall	735	36.12	19.04	70.09	21.95