Beliefs and Endogenous Cognitive Levels: an Experimental Study*

Marina Agranov[†], Elizabeth Potamites[‡], Andrew Schotter[§] and Chloe Tergiman[¶]

October 2011

Abstract

In this paper we use a laboratory setting to manipulate our subjects' beliefs about the cognitive levels of the players they are playing against. We show that in the context of the $\frac{2}{3}^{\text{rds}}$ guessing game, individuals' choices crucially depend on their beliefs about others' levels. Hence, a subject's true cognitive level may be different than the one he exhibits in a game with the difference being attributed to his expectations about the sophistication of the players he is matched with.

*We thank the Center for Experimental Social Sciences (CESS) at New York University for providing the financial support for this study. We also thank Colin Camerer, Giorgio Coricelli, Vincent Crawford, Nagore Iberri, Rosemarie Nagel and Stefan Penczynski for very helpful comments.

†Caltech and CESS.

‡Mathematica.

§NYU and CESS.

¶UBC and CESS.

1 Introduction

Models positing a homogeneous population of rational agents often fail to explain data generated by controlled laboratory experiments. This point has been made in dramatic fashion by Nagel (1995), Stahl and Wilson (1995), Ho, Camerer and Weigelt (1998), Bosch-Domenech, Montalvo, Nagel and Satorra (2002), Costa-Gomes and Crawford (2006), and many others following their lead (for an up to date survey of the literature see Crawford, Costa-Gomes and Iriberri (2010)). These models assume the existence of a set of players that are heterogeneous in their levels of rationality.

In this paper we view a person's choice (which often interpreted as his cognitive level) as endogenously determined by his expectations about the (also endogenously determined) cognitive levels of other players, which in turn affect their choices.[1] We use a controlled laboratory experiment to show that observed level of play is a function of how sophisticated a player believes his opponents to be. We study the $\frac{2}{3}^{\text{rds}}$ guessing game, as discussed by Nagel (1995). In this game, a cognitive hierarchy model typically assumes that level zero agents choose randomly. Level one agents best respond to level zeros. More generally, agents who are level k are either assumed to believe all other players are of level $k-1$ and best respond to them (as in Nagel (1995) and Stahl and Wilson (1995)), or are assumed to best respond to a belief on the distribution of lower levels in the population (as in Camerer, Ho and Chong (2004)). Therefore, the choice of a participant in the $\frac{2}{3}^{\text{rds}}$ guessing game, which we can observe and call the "observed cognitive level," depends on the following three elements: (1) the subject's beliefs about level zero play; (2) the subject's expectation about the cognitive levels of other players; and (3) the number of steps of reasoning that the subject is capable of in the context of the game.

This second element (the beliefs about cognitive levels of other players) is a crucial component that determines subjects' choices in the $\frac{2}{3}^{\text{rds}}$ guessing game. Indeed, it is plausible

[1] Theoretically, the only model that we are aware of that has endogenous choice of cognitive levels is Choi (2011) who posits that agents differ by the cognitive costs they have in functioning at different rationality levels. As a result, while agents cannot act as if they were smarter than they actually are, they can choose to behave in various ways depending on their comparison between the marginal benefit and marginal cost of increasing their rationality level.

that subject of a high cognitive level may choose an action that is associated with players of lower cognitive levels if he believes that others in the group are of much lower cognitive levels compared with his own.[2]

The goal of this paper is to illustrate that the distribution of observed cognitive levels crucially depends on the beliefs of our subjects regarding the sophistication of their opponents. To achieve this goal we use a number of manipulations regarding the expectations about the cognitive levels of other players. We start by having undergraduate students play the $\frac{2}{3}^{\text{rds}}$ guessing game against a group of graduate students who have training in these sorts of games ("Graduate" treatment). We find that relative to the "Control" treatment, in which undergraduate students are playing against each other, there is a significant shift in the distribution of cognitive levels towards more sophisticated behavior when undergraduate students are playing against graduate students. We then ask (different) students to play against a set of computers who play uniformly on the support. The subjects are aware of the strategy the computers follow. In this treatment (the "Computer" treatment) students are now essentially playing only against level zero agents. Relative to the Control treatment, we observe a shift towards less sophisticated behavior.[3]

As a by-product of our design, we are able to accomplish a second goal which is to measure, in a clean and precise manner, the fraction of people who are capable of at least one step of reasoning (strategic players). Indeed, in all other papers the identification of strategic types (i.e., types above zero) is confounded by the fact that the behavior of subjects is a function of both their objective cognitive levels and their beliefs about others. This is in

[2] The idea that one's behavior depends on both his own cognitive level as well as his expectations on the cognitive levels of others is explored theoretically by Strzalecki (2010) who studies the Email Game and shows that coordination can be achieved in a finite number of steps.

[3] Although the focus is different, there are a limited number of papers that we are aware of that control for the beliefs of the subjects in some way. Coricelli and Nagel (2009) use fMRI to measure brain activity when subjects participated in the $\frac{2}{3}^{\text{rds}}$ guessing game in order to differentiate between subjects making random choices and those making choices resulting from higher levels of reasoning. Our results align with theirs in that some subjects differentiated their behavior in the human compared with computer treatment, in which computers chose numbers randomly. Costa-Gomes and Crawford (2006) study the two-person guessing game in which they train subjects subjects and show that when these trained subjects are paired with computers playing a known strategy they exhibit behavior consistent with Level-k theory. Using the two-person centipede game with repeated rounds, Palacios-Huerta and Volij (2011) show that when players of differing cognitive abilities are faced with each other, they converge to the equilibrium at different rates.

sharp contrast with our Computer Treatment where we completely control for these beliefs: all subjects capable of at least one step of reasoning should behave as level 1 players. Thus, we can clearly identify those who are capable of at least one step of reasoning. While the fraction of subjects capable of at least one step of reasoning is stable across treatments (about 50%), as we move across treatments from the Computer to the Control and finally to the Graduate treatment, we observe a redistribution of types to higher levels. In other words, as we move across treatments, our results suggest that some of those acting as level 1 reasoners in the Computer treatment would behave as level 2s or more in the Control or Graduate treatment.

The Control, Graduate and Computer treatments use a between subjects design. To assess the robustness of our results regarding the manipulation of subjects' beliefs, we also conduct a experiment with a within-subject design ("Combo" treatment). In the Combo treatment, we use the strategy method to obtain subjects' choices in 8 different $\frac{2}{3}^{\text{rds}}$ guessing games. In each game the undergraduate students are playing against a group which consists of a combination of computers and graduate students. The groups range from 7 computers and 0 graduate students ($(7\text{c}, 0\text{g})$ configuration) as in the Computer treatment, to 0 computers and 7 graduate students ($(0\text{c}, 7\text{g})$ configuration), as in the Graduate treatment. We start by showing that the choices from the $(7\text{c}, 0\text{g})$ configuration in the Combo treatment match the choices from our Computer treatment. There is also no statistical difference between the answers from the $(0\text{c}, 7\text{g})$ configuration in the Combo treatment and the choices in the Graduate student treatment. This suggests that the strategy method did not have an impact on the behavior of the subjects. This treatment further allows us to observe and follow the behavior of the same subjects when their beliefs about the population they are playing against change. Finally, because subjects submit all their choices at the same time, their choices in the different configurations of the group are not influenced by experience in the $\frac{2}{3}^{\text{rds}}$ game. While a little more than half of the population seems to play randomly, the rest climb the cognitive hierarchy ladder because they believe that others are climbing it as well and they must keep up with them (best respond).

Our experimental design provides an interesting stress test of the various theories of behavior that have been used to explain choices in the $\frac{2}{3}^{\text{rds}}$ guessing game. As we discuss above, Level-k and Cognitive hierarchy models are two main theories that has been developed to interpret the data in the guessing games. One of the main features of both theories is the idea that subjects do not contemplate people who are of a higher cognitive levels than they are. Thus, our Graduate treatment places the subjects in an "out of model" situation, and raises the question of how will an undergraduate subject who adhere to the assumptions of one of these models behave when playing against set of graduate students. We discuss the implications of those theories for the behavior in our experimental setup. We reach the conclusion that one need not abandon Level-k and Cognitive hierarchy models to explain behavior in experiments like ours. However, once the information about the sophistication of one's opponents is injected in these models (information which is outside of these models), one need to make additional assumptions regarding the level of sophistication of the most sophisticated players to generate predictions of these models.[4] More generally, it turns out that any model of bounded rationality which satisfies two plausible assumptions (belief monotonicity and best responding) will predict that the distribution of observed choices shifts monotonically to the left as the population of opponents faced by a subject becomes more sophisticated (exactly what happens in the Combo treatment when the group moves from being composed of all computers to all graduate students).

Our paper contributes to the literature that investigates the behavior of individuals in strategic one-shot games, using the guessing game as a case study. Recent research has lead to several breakthroughs in understanding the discrepancy between predicted equilibrium behavior and behavior observed in experiments. For example, Costa-Gomes and Crawford (2006) conducted a series of 16 two-person guessing games using MouseLab to show that for a significant fraction of the population, deviations from equilibrium can be attributed to how subjects model others' decisions, and that level-k models can explain those responses. Burchardi and Penczynski (2010) design an experiment in which players are divided into teams. Each member is allowed to pass both his/her individually preferred choice and a

[4] We show one such assumption that does the job. Assuming that undergraduate students that know how to behave in a level-k model believe that all graduate students are one level above them, we can pin down the behavior of undergraduate students in both Graduate treatment and Combo treatment.

persuasive message to his/her partner concerning how best to play the game. The authors then classify these verbal arguments according to levels of strategic sophistication. Agranov, Caplin and Tergiman (2010) introduce a novel experimental technique that allows them to record the intermediate choices of subjects in the $\frac{2}{3}^{\text{rds}}$ guessing game in the three minute period immediately after the structure of the game has been conveyed to them. The authors find that average choices decrease over time indicating an increase in strategic sophistication, and that level 0 behavior matches well the standard assumptions of level-k models. Coricelli and Nagel (2009) use fMRI to study the differences in the neurological responses of players with different levels of strategic sophistication. Grosskopf and Nagel (2008) focus on bounded rationality to explain the failure to play the equilibrium strategy in the $\frac{2}{3}^{\text{rds}}$ guessing game. The authors use a two-person $\frac{2}{3}^{\text{rds}}$ guessing game in which the weakly dominant strategy is to play zero regardless of the belief one holds about the choice of the other player. The authors found that even in this circumstance, in early rounds of the game, the majority of people choose numbers above zero.

Separate studies have shown that choices in the $\frac{2}{3}^{\text{rds}}$ guessing game vary either with repeated play or with subject pools. For example, Weber (2003) has subjects play a series of $\frac{2}{3}^{\text{rds}}$ guessing games without feedback. The author shows that subjects' choices decrease over time. Camerer, Ho and Chong (2004) summarize a series of experiments that use separate pools of subjects (CEO's, portfolio managers, Caltech board, game theorists, university students and high school students) and show that choices in the guessing game vary with these subject pools. However, in these papers, subjects' choices can be influenced by several factors. Indeed, in Weber (2003) the change of choices can come from learning the structure of the game or the belief by players that other players' choices will go down. In the experiments that Camerer, Ho and Chong (2004) survey, the difference in the distribution of choices can come from the difference in a subject's own objective cognitive level or his/her beliefs about the objective levels of his/her opponents. Neither of these previous studies highlight the role of beliefs, which is the main point of our paper.

Finally, while we are not aware of any paper that directly addresses the question of how beliefs affect choices in the $\frac{2}{3}^{\text{rds}}$ guessing game, there are three recent papers that investigate to some extent related questions. Kocher et al (2009) studies naive advice and observation of others' decisions in the $\frac{2}{3}^{\text{rds}}$ guessing game and finds that advice triggers faster convergence to equilibrium. Georganas et al (2011), among other things, let players play against three different types of opponents: randomly selected one from the population, the opponent who scored highest on the strategic intelligence quiz and the opponent who scored lowest. Authors fund that some players adjust their choices when playing with stronger opponents, but neither quiz scores nor levels predict which subjects make this adjustment. Buhren et al (2010) conducts $\frac{2}{3}^{\text{rds}}$ guessing game among chess players and find results very similar to those obtained when general population is used. Authors attempt to manipulate beliefs of the participants, however, players are not randomized into control groups, and when "playing" against opponents of different skills the participants have already played the game at least once.[5]

The rest of the paper is organized as follows. In Section 2 we describe the experimental procedure and design. In Section 3 we discuss the relationship between our experimental design and various behavioral models (level-k, cognitive hierarchy and iterated elimination of dominant strategies) and formulate hypotheses with reference to these theories. In Section 4 we analyze the experimental results. The conclusions are in section 5.

2 Experimental Design

The experiments were conducted in classrooms of undergraduate students at New York University and at the University of British Columbia.[6] In total, 329 students participated in

[5]The part in which chess players play against opponents of different skills is implemented using an online survey. Chess players first participate in a standard beauty contest game. After having found out the target number those players who agreed to participate in a second experiment were asked to give a guess what the target number in experiment 1 had been for players of different ratings. The authors find a weakly downward sloping trend.

[6]The UBC students played in the Combo treatment. The NYU students played the Control, Graduate and Computer treatments. The NYU students were taken from various first year classrooms. While we determined ahead of time which treatment we were going to run, at NYU, classes are not organized by GPA and so there is no reason to suspect that they differed in GPA, which could drive the results rather than beliefs. The treatment at UBC also provides strong support for this since the subjects there, in the zero graduate students and 7 graduate students matched the behavior of the NYU students in the Computer and Graduate student treatments, respectively. See footnote 27 in Result 5 of Section 4 for evidence that the

these experiments. We conducted four treatments: the Control treatment (91 participants), the Graduate treatment (99 participants), the Computer treatment (85 participants) and the Combo treatment (54 participants). The experimental procedure was as follows. At the beginning of a class lecture, we distributed the instructions of the experiment, read those instructions out loud and asked students to write their number (or numbers for the Combo treatment) on a piece of paper, which they put in an envelope. We collected the envelopes and calculated the payoffs of the participants while they were attending the rest of their lecture. At the end of it, we distributed the envelopes back to the students: those who won in this experiment had $10 cash in their envelopes, the other participants had nothing. That way we ensured that the identity of the winners remained anonymous. In the Combo treatment students submitted several numbers, each corresponding to their choice for a different configuration of the group they were playing with. We then drew at random one configuration and paid them according to the number they chose for that case.

All the experiments lasted less than 10 minutes in total, including reading the instructions.

In the Control treatment, our subjects played against each other in groups of size 8. We compare the data from the Control treatment to the other experiments reported in the literature. They also serve as the baseline for comparison with the Graduate and Computer treatments. Subjects were given the following instructions:

> "Choose a number between 0 and 100. You will be put into groups of 8 people. The winner is the person whose number is closest to $\frac{2}{3}$ times the average of all chosen numbers of the people in your group. The winner gets a fixed prize of $10. In case of a tie the prize is split among those who tie."

Our first manipulation (the Graduate treatment) aims to put participants in the situation, in which they play the game against a group of very sophisticated players. To achieve this we first conducted a $\frac{2}{3}^{\text{rds}}$ guessing game with a group of 8 graduate students from the

behavior of NYU and UBC students did not significantly differ.

department of economics at NYU using the instructions from our Control treatment. We then conducted a $\frac{2}{3}^{\text{rds}}$ guessing game experiment in which undergraduate students were playing against 7 of the 8 graduate student that completed the game before (so that that the group size remained at 8). The undergraduate subjects received the following instructions:

> "Choose a number between 0 and 100. You will win $10 if your chosen number is closest to two thirds times the average of all chosen numbers of the people in your group.
>
> Your group: 8 graduate students in the Department of Economics, who have training in these types of games, played this game a few days ago. You will replace one of them. So your group is YOU and 7 of those graduate students.
>
> You will win $10 if your chosen number is closest to $\frac{2}{3}$ times the average of all chosen numbers (yours and 7 graduate students). In case of a tie the prize is split. Notice you are not playing against people in this room. Each of you is playing against 7 graduate students. So, all of you may earn $10 and none of you may."

The goal of our second manipulation (the Computer treatment) was to shift the beliefs about the other players in the opposite direction. In this treatment, an independent group of undergraduate students played against seven computers who chose randomly between 0 and 100. The instructions for this experiment were:

> "Choose a number between 0 and 100. You will win $10 if your chosen number is closest to $\frac{2}{3}$ times the average of all chosen numbers of the people in your group.
>
> Your group: Your group consists of you and 7 computers. Each of those computers will choose a random number between 0 and 100, each number being equally likely. So your group is YOU and 7 computers.
>
> You will win $10 if your chosen number is closest to $\frac{2}{3}$ times the average of the numbers in your group (yours and the 7 random numbers chosen by the computers). Notice you are not playing against people in this room. Each of you is playing against 7 computers. So, all of you may earn $10 and none of you may."

Computers are essentially playing a level zero strategy. In this treatment we expect all subjects who are capable of at least one level of reasoning to choose numbers close to 32.[7] Subjects who are capable of higher steps of reasoning will appear to be level one. Further, this treatment provides us with a way to estimate how many participants are capable of at least one level of reasoning, which we will be able to compare with the proportions in the Control and the Graduate treatments.[8]

Finally, we ran the Combo treatment in which we used the strategy method to elicit our subjects' choices for mixed groups. For each undergraduate subject, the group consisted of himself, X computers and $7 - X$ graduate students randomly taken from the set of eight graduate students that played the game before, where X took the values 0, 1, 2, 3, 4, 5, 6 and 7.[9] In other words, instead of facing seven computers (as in the Computer treatment) or seven graduate students (as in the Graduate treatment), students in the Combo treatment faced a mixture of those two populations. For each specified parameter X, we elicited the choices of the subjects.[10] At the end of the experiment, we randomly chose one of the configurations and students were paid based on the numbers they chose in that configuration as well as the numbers chosen by the computers and the graduate students.[11] This treatment allows us to establish the connection between a subject's observed cognitive level and his beliefs about the cognitive level of his opponents, as those beliefs change from them having a low cognitive level (as in the Computer treatment) to a them having a high cognitive level (as in the Graduate treatment). Importantly, because we use a within subjects design, we can trace the behavior of the *same* participant. Further, because subjects submit all their answers at the same time, our subjects' choices are not impacted by their experience.[12]

[7]The best response to an average of $\bar{x}$ if a subject's number is counted in the average is $\frac{2(n-1)\bar{x}}{3n-2}$ where n is the total number of players in the group. In our case $n = 8$ and thus subject should choose 31.8. However, if a subject fail to recognize that his own number influences the average of the group, he may choose 33.33.

[8]Notice that even though the Computer treatment is a risky environment, the degree of risk-aversion should not affect the behavior of a player in this treatment. EXPAND WHY.

[9]Notice that in all treatments, each participant is playing the game against the group of 7 subjects that played the game in the past. Thus, the number of participants in each treatment does not need to be divisible by 8, as subjects in the classrooms do not interact with each other.

[10]See Appendix A for the complete instructions for the Combo treatment.

[11]For example, if there were 3 graduate students in the configuration, we randomly chose 3 numbers from those submitted by the graduate students.

[12]We know that repeated play, with feedback (see Nagel (1995)) or without feedback (see Weber (2003)) can lead subjects to choose lower numbers over time.

3 Theoretical predictions

In this section of the paper we will discuss the relationship between our experimental design and various theories of behavior that have been or could be used to explain behavior in the $\frac{2}{3}^{\text{rds}}$ guessing game.

Before we start, it is important to point out that our experimental design stretches the assumptions of most standard theories governing behavior in the $\frac{2}{3}^{\text{rds}}$ guessing game by placing subjects in situations not governed by any of these theories. More specifically, we consider the behavior of subjects who are endowed with a variety cognitive levels, k, each representing the limit of their cognitive abilities and who play initially against a set of graduate students described to them as experienced in the game being played. These opponents can be assumed to be of a higher level of cognition than any of our subjects (although some may feel otherwise) and by announcing who they are playing against we tried to induce that belief in our subjects. This fact places our subjects in an "out of model" situation since in all standard theories (level-k, cognitive hierarchy, or simple best response to beliefs) subjects do not contemplate people who are of a higher cognitive levels then they are. Our subjects are asked to choose to behave in this unorthodox (from the standpoint of theory) situation.

In fact our graduate students are also in an out-of-model situation since they are playing against a set of subjects described to them as being of their own level so that violates the assumption that in a level k model all subjects believe that everyone else is one level below them. However, that does not stop any graduate student from believing that all of his cohorts are one cognitive level below him since they were only told the others are graduate students and not that they were all intellectual clones. This assumption gives them something to best respond to. If they did not assume that, we could close the model by assuming that our graduate students were all smart enough to know the concept of iterative deletion of dominated strategies, assume that this fact was common knowledge among them, and then deduce the fact that they would all choose zero. In other words, in level-k models where subjects are led to believe that all of their opponents are of an equal cognitive level, some additional assumption must be added to the model to identify behavior.

In our experiment, the question raised is how will an undergraduate subject whose thinking is consistent with a level-k model play when facing the set of graduate students described above? If our undergraduate subject is of level k and adheres to the assumptions of that theory, he will know that those graduate students will be best responding to some subjects whose level of cognition is just below them. So if graduate students are in fact level $j > k$, then graduate students will best respond to the assumption that all other subjects are of level $j - 1$. The question is how far above them do our undergraduate subjects think these graduate students are?

To make things tractable, we will assume that our subjects, knowing how to behave in a level-k model, make the assumption that all graduate students are one level above them. In other words, we amend level-k theory and add the possibility that someone who would be called level k in the standard theory, assumes that all people suspected to being of a higher cognitive level is exactly one level higher than they are. Once this is assumed, our level k subject is perfectly capable of best responding to that belief and choosing appropriately. With that assumption we can pin down behavior for any subject since, knowing how to best respond, our subjects can calculate that all graduate students will choose $\frac{2}{3}^{\text{rds}}$ of their contemplated choice and hence choose $\frac{2}{3}^{\text{rds}}$ of that. For example, say a subject in our experiment was of level 2 and would play 22 against other undergraduate students. When faced with graduate students, our level 2 undergraduate assumes they are of level 3. Since a level 2 chooses 22 (under the assumption of a random level zero) he will think that all graduate students will choose a best response to 22 or 14. Our subjects will then choose 9.5. So, for our assumed level 2 subject, his choice moves from 22 to 9.5 under the assumption that he is acting against a set of 7 graduate students. Since our undergraduates are of varying levels of cognition, this analysis will lead each of them choose differently (a level 1 will wind up choosing 14, for example) so we should still observe a distribution of actions by our subjects but it should be shifted to the left when compared to our Control Treatment. Only level 0 players would not shift as they are non-strategic.

It is interesting to note here that under the above argument we actually have subjects choosing at levels above their presumed cognitive limits (in strict interpretation of level-k theory) since in the example above we assumed that our level 2 subjects started out choosing 22 but then decreased this choice to 9.5 in response to being told that his opponents were more sophisticated than him. This is not as strange as it might seem since all they are doing, after assuming that those above them are just one level above, is applying one iteration of best responding, a capability totally within their grasp.

So, while a level-k model or a cognitive hierarchy model assumes that no agent in the model thinks that any agent exists who is a higher level than him, when the agent is informed about the existence of such agents, they are able to best respond as long as they can understand the type of behavior such higher level subjects are employing. The assumption made above, that all undergraduates assume that all graduate students are one cognitive level above them, allows our undergraduate to behavior in a way that is within their strategic capability. It is an assumption made to allow us to identify behavior for any level-k undergraduate and is comparable to the identifying assumption in standard level-k models about the zero-level types. In other words, when the model assumes that all agents assume that others are below them in cognitive level, we need an identifying assumption about the lowest level agent, while when we assume that all other agents are above you, we need an identifying assumption for the highest cognitive-level agent. Having described behavior of our subjects when they face a set of opponents composed solely of graduate students we can easily adjust how they will behave when we mix computers with them since each computer is assumed to choose 50 and so it is easy adjust a subject's best respond for any mixture of computers and graduate students, again under the assumption that all graduate students are one level above them.

This way of thinking expands our expectations about behavior since it now seems possible that in response to their beliefs about their fellow opponents, subjects are able to change their behavior in either direction, i.e., look as if they are more or less sophisticated than they actually are. This is still consistent with the punch line of our paper, however, since all we are claiming here is that observed behavior in games like the $\frac{2}{3}^{\text{rds}}$ guessing game are a function of the beliefs of the agents about the sophistication of those they face.

When it comes to cognitive hierarchy models we again place our subjects in an out-of-model situation since we define the hierarchy for them rather than assume that it is of any particular type, such as Poisson. Here the problem for our subject is similar to that of the level-k subjects discussed above since, while it is easy for them to choose their behavior when faced with a set of computers, they too need some help in deciding how sophisticated graduate student will behave. Again, the existence of graduate student violates the assumption that each subject believes he is at the top of the hierarchy looking down. Still it should be obvious that whatever assumption one ultimately makes about these graduate students, the higher their proportion, the lower the mean choice, and hence the lower our subjects' best response.

To generalize the argument above, consider any model of bounded rationality which satisfies the following two assumptions:

- *Assumption I "Belief Monotonicity"*: If an agent added to a population of agents playing the $\frac{2}{3}^{\text{rds}}$ guessing game is at a level at least as high the maximum level currently existing, then the probability distribution defining choices in the population shifts to the left and the mean of the population decreases.
- *Assumption II "Best Responding"*: The model of bounded rationality we are looking at requires that all agents best respond to their beliefs.

Assumption II alone (without the Belief Monotonicity assumption) predicts very specific behavior in the computer treatment (as well as the combo treatment with no graduate students) and leads to our first hypothesis.

Hypothesis 1: Subjects who adherent to either a level-k model of behavior or a cognitive hierarchy theory and have level of sophistication of at least one should choose 31.8 in the Computer treatment.

This is true because all subjects whose cognitive level is above level 0, knowing how to best respond, will easily see that 31.8 is a dominant action in the all computer situation.

They will all behave as if they were level 1 even if they were of a higher level.

Taken together, Assumptions I and II imply that as we add more graduate students in the combo treatment (substitute computers for graduate students) we get a shift in the distribution of choices to the left.

Proposition 1: In a level-k model (or a model of cognitive hierarchy) as we add more and more graduate students to a population of computers playing the $\frac{2}{3}^{\text{rds}}$ guessing game against our undergraduate subject, the distribution of choices made by our undergraduate subjects shifts to the left.

Proof: The proof follows easily from Assumptions 1 and 2 and the identifying assumption made above about how graduate students are assumed to play. When there are only computers all subjects, no matter what their level, assume that they face a level-zero machine with a mean of 50 and hence best respond to that. If we then introduce our first graduate student, under the identifying assumption made above, all subjects must assume that the graduate student will choose less than 50 and hence the mean will fall. What level of choice the graduate student is assumed to make depends on the level of cognition of the undergraduate student; the higher his level the lower he will assume the graduate student choice is but, by construction, the graduate student choice must be less than 50. Hence each undergraduate will shift his choice to the left (unless he is level zero). The same occurs as we add more and more graduate students, **q.e.d.**

The import of Proposition 1 is that one does not need to abandon the types of bounded rationality models used to explain behavior in experiments like ours. However, our experiment does stress test these models by injecting information about the sophistication of one's opponents that is outside of the model and seeing how the model responds is interesting. What one cannot do is have faith that the behavior one observes in experiments or elsewhere is an honest indicator of the true distribution of cognitive types. What we observe is contaminated by the beliefs of our subjects.

Hypothesis 2 then follows.

Hypothesis 2: As the population of opponents faced by any subject in our experiment become more sophisticated, (i.e., moves from being composed of all computers to all graduate students) the distribution of choices observed in our experiment should shift to the left.

Note that the analysis above is tailor-made to explain behavior in our experiment since in the experiment we have structured the beliefs of our subjects by telling them the exact composition of the types of subjects that they will be facing. In more general settings, where such a structure is not provided, the situation is more difficult. When economic agents face a generic strategic situation they must make an assessment of the distribution of sophistication of their opponents. However, this assessment just represents their first order beliefs. They must then contemplate what their opponents believe about the sophistication of others and what their opponents belief about what they believe about the sophistications others etc. This can get complicated but however complicated one makes these models, it should still be obvious that in any such model, if we reveal information that indicates that subjects in the experiment are either more or less sophisticated than previously expected, the observed set of choices will change appropriately.

In summary, while level-k-type models can be used to generate the type of comparative static results observed in our data our results makes the task of inferring a subjects true cognitive type from his or her choices more complicated because such choice data does not control for the beliefs that subjects hold about each other or their higher order beliefs. A comprehensive analysis of choice in such games then might have to jointly estimate beliefs and cognitive levels - a complicated task.

4 Results

We structure the results around the hypotheses presented in Section 3. We start by showing the similarity between the Control treatment and the behavior reported in other experimental studies of the $\frac{2}{3}^{\text{rds}}$ guessing game (Result 0). We then present the data regarding Hypothesis 1, which suggests that subjects who have level of sophistication of at least one should choose 31.8 in the Computer treatment. This hypothesis allows directly to measure the fraction of the population that is not capable of strategic thinking and is usually referred to as level-0 according to the terminology of level-k and cognitive hierarchy models (Result 1). After that, we proceed to Hypothesis 2, which suggests that any model of bounded rationality that satisfies Belief Monotonicity and Best Responding assumptions should exhibit the monotonicity with respect to sophistication. We evaluate this hypothesis in two steps. First, we show that distribution of choices in the Graduate student treatment is shifted to the left relative to the distribution of choices in the Control treatment, while the opposite is true in the Computer treatment. This suggests that when playing against 7 graduate students, undergraduate students respond by choosing lower numbers, and when playing against 7 computers they choose relatively higher numbers (Result 2a). Second, we look at the Combo treatment, which provides a more refined test of Hypothesis 2, and ask whether the increase in the number of graduate students in the group has a monotonic effect on the average choice of undergraduate students (Result 2b). Result 2b investigates the monotonicity of aggregate choice. However, the Combo treatment allows us to track individual behavior and look at how these aggregate results are generated (Result 3). Finally, we show that the strategy method used in the Combo treatment does not interfere with strategic thinking of subjects (Result 4).

Result 0: Behavior in the Control treatment is similar to the behavior reported in other experimental studies of the 2/3 guessing game.

We start by comparing the results of our Control treatment with the results of other experiments reported in the literature. Table 1 shows some descriptive statistics from the Control treatment and other studies. In Table 2 we report the distribution of cognitive levels in the Control treatment according to Nagel (1995)[13] and compare this distribution to

[13]Nagel's classification starts from the premise that Level 0 players choose 50. Nagel then constructs neighborhood intervals of $50p^n$, where p is the multiplier used in the game ($p = \frac{2}{3}$ in our case) and n represents the level of reasoning ($n = 0, 1, 2, ...$). The numbers that fall between two neighborhood intervals of $50p^{n+1}$ and $50p^n$ are called interim intervals. To determine the boundaries of adjacent intervals a geometric mean is used. Thus the neighborhood interval of $50p^n$ have boundaries of $50p^{n+\frac{1}{4}}$ and $50p^{n-\frac{1}{4}}$ rounded to the nearest integers. The exception is Level 0, which is truncated at 50. Nagel classifies as Level 0 choices between 45 and 50, Level 1 those between 30 and 37, Level 2 between 20 and 25, and Level 3 those between

Nagel's data.

	mean choice	median choice	st dev	group size
Control treatment	35.1	33	21	8
Nagel (1995)	37.2	33	20	$14-16$
Ho, Camerer and Weigelt (1998)	38.9	NA	24.7	7
Agranov, Caplin and Tergiman (2010)	36.4	33	20.2	8

Table 1: Summary Statistics for the Control treatment.

	Control Treatment	Nagel's data
Level 0	8%	7.5%
Level 1	25%	26%
Level 2	18%	24%
Level 3	8%	2%
Fraction captured by Nagel's classification	59%	59.5%

Table 2: Level classification according to Nagel (1995).

It is clear from Tables 1 and 2 that the subjects in the Control treatment are very similar to other studies in terms of the their chosen numbers as well as the distribution of cognitive levels.

Result 1: Evidence from the Computer treatment suggests that only about one half of subjects are capable of at least one level of reasoning.

The Computer treatment provides a clean measure of the fraction of the population that are capable of best responding, which is what strategic subjects necessarily can do according to level-k and cognitive hierarchy models. In other words, in the Computer treatment any subject with a cognitive level of at least 1 should have an observed cognitive level of exactly 1 and should choose 31.8, since 31.8 is the best response to the computers choosing randomly. Notice that a subject may be capable of one or more levels of reasoning, but may fail to realize that the number he is choosing affects the average of the group. In this case, he/she should choose 33.33. Since our interest is in estimating the fraction of strategic players and

13 and 16.

not in how many people realize that their chosen number counts in the average of the group, we will stick to Nagel (1995) classification and consider all numbers between 30 and 37 as the level 1 answers in the Computer treatment. Using this definition means that we implicitly allow subjects to make small mistakes in calculating their answers.

There are 49% of subjects that chose numbers between 30 and 37 in the Computer treatment.[14] This means that the remaining 51% of subjects are not capable of best-responding, which is the necessary condition for being strategic in this environment. Put differently, we observe about one half of the population behaving non-strategically (as level 0) in our experiment. Such high percentage of level 0 subjects may seem surprising at first. However, if one looks at the Control and Graduate students treatment and uses Nagel's classification to infer the percentage of the strategic players the results are similar to what we obtain in the Computer treatment. Indeed, 55% of subjects can be classified as being level 1 or more in the Graduate treatment and 51% in Control treatment. A test of proportions indicates that these percentages are not different from each other and also not different from the 49% in the Computer treatment (all pairwise p-values are greater than 10%).

Our results are in line with the recent experimental work that investigates whether people are able to best-respond in various strategic situations. Several papers use $\frac{2}{3}^{\text{rds}}$ guessing game for this purpose and find that significant fraction of subjects behave in a non-strategic way. Agranov, Caplin and Tergiman (2011) use a new method that allows for the tracking of choices over time by each subject and document that almost 60% of subjects appear to make random choices. Buchardi and Penszynski (2010) analyzed subjects arguments while attempting to convince their "teammates" to follow their advice and find that the fraction of level 0 players is about one third. Ivanov, Levin and Niederle (2010) study the "maximal game", a variant of the standard second-price common value auction, and show that the winner's curse is unlikely to be driven by beliefs. The authors document that among those subjects who overbid (roughly 40%) a large majority (about 80%) fail to best respond to

[14] The breakdown of choices in that interval is: 5 subjects chose 30, 1 subject chose 32, 23 subjects chose 33, 2 subjects chose 33.3, 7 subjects chose 34, 3 subjects chose 35 and 1 subject chose 37. Interestingly, 33 (the modal choice in that interval and the data in its entirety) is the best response to computers playing randomly if one's choice isn't counted in the average.

their *own* past choices. This parallels our finding that in the Computer treatment, in which beliefs of subjects regarding their opponents are controlled, the deviations from theory are unlikely to be driven by beliefs and, thus, failure to behave as level 1 indicates the failure to best respond, which is the essence of the strategic behavior.[15]

Is it surprising that some people fail to best-respond? The answer must depend on the nature of the game in question. It would be unreasonable to expect all subjects to behave optimally and best respond to their beliefs about others in a complicated situation (like chess, for example). On the other hand, if a game is simple enough and best-response is rather obvious we would expect large proportion of subjects to realize that and play it. Recent paper by Dufwenberg, Sundaram and Butler (2010) makes this point. The paper uses the "game of 21" (variant of Race game[16] also studied by Gneezy, Rustichini and Vostroknutov (2010)) to investigate whether players are able to find a dominant strategy. Authors emphasize that some subjects experience a moment of "epiphany", which is sudden realization that the game they are playing has an analytical solution (epiphany 1). This realization is the first step towards figuring out what is the solution of the game (epiphany 2). The paper documents that those people that play a simpler version of the game of 21 (game of 6) first and then move on to the game of 21 are more likely to have epiphany 1 and thus perform better in the more complicated game. In other words, while some participants were able to figure out the dominant strategy in the complicated game almost right away, others could do so only after playing the simpler version of the game analytical solution of which is similar to a more complicated game but involves less steps of backward induction reasoning. Our results suggest that $\frac{2}{3}^{\text{rds}}$ guessing game in which undergraduate students play against the group of computers is simple enough that half of the population is able to figure out optimal action and, at the same time, is complicated enough so that another half finds it hard to do so.

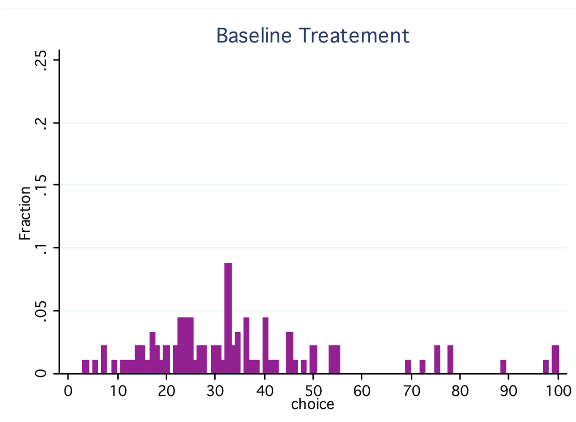
Our main hypothesis in this paper is that the observed distribution of cognitive levels

[15]redIn a different setup (Tennis Coach Problem) Arad (2011) finds that about 20% of participants behave in a non-strategic way.

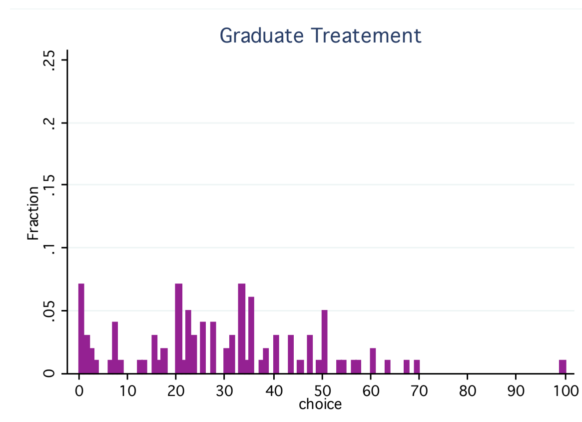
[16]Race game is a two-person zero-sum sequential game with perfect information. At each stage, two players alternate who is the mover. The mover can choose any number from the pre-specified set of numbers. The player whose chosen number brings the total sum of all chosen number so far to particular mark (21 in the "game of 21") wins the game.

is endogenous and can be manipulated by varying the expectations about other players' cognitive levels. More specifically, as we have established in the theory section, any model of bounded rationality that satisfies Belief Monotonicity and Best Responding assumptions would suggest that the choices of undergraduates are monotonic with respect to the sophistication of the players they are playing against, with graduate students being the most sophisticated and computers being the least sophisticated.

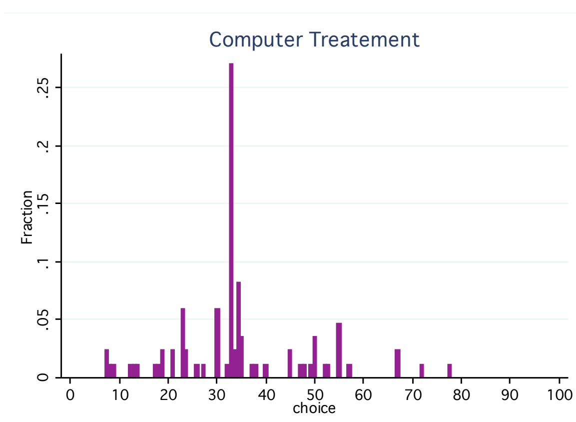
Result 2a: **The distribution of observed cognitive levels in the Graduate treatment is shifted towards higher cognitive levels compared with the Control treatment. The opposite is true in the Computer treatment.**



(a) Control



(b) Graduate



(c) Computer

Figure 1: Histograms of choices in the Control, Graduate and Computer Treatments.

In Figure 1 we present the histograms of choices in the Graduate and Computer treatments. In Tables 3 and 4 we present summary statistics of the choices observed in these two treatments, and the distribution of cognitive levels in these treatments according to the Nagel (1995) classification, respectively.

	Mean Choice	Median Choice	Std. dev.	# Obs
Control Treatment	35.1	33	21.02	91
Graduate Treatment	28.6	27	18.93	99
Computer Treatment	34.3	33	14.07	85

Table 3: Summary Statistics of the Control, Graduate and Computer treatments.

	Control	Graduate	Computer
Level 0	8%	10%	9%
Level 1	25%	20%	49%
Level 2	18%	20%	
Level 3	8%	5%	
Level ∞	0%	10%	
Fraction captured by Nagel's classification	59%	65%	58%
Fraction not classified	41%	35%	42%

Table 4: Level classification of Nagel (1995) in the Control, Graduate and Computer treatments.

While the percentage of subjects who are classified as at least level 1 is similar across these treatments (as we established in Result 1), the distribution of types within these groups changes as we match our subjects with more sophisticated opponents.[17] As a result, the distribution of choices across treatments changes in the expected direction.

[17] As we can see from Figure 1, the distribution of choices in the Graduate treatment is shifted to the left relative to the one in the Control treatment. Note, for instance, that there are no subjects that chose numbers below 1 in the Control treatment, while 10% of the population did so in the Graduate treatment.

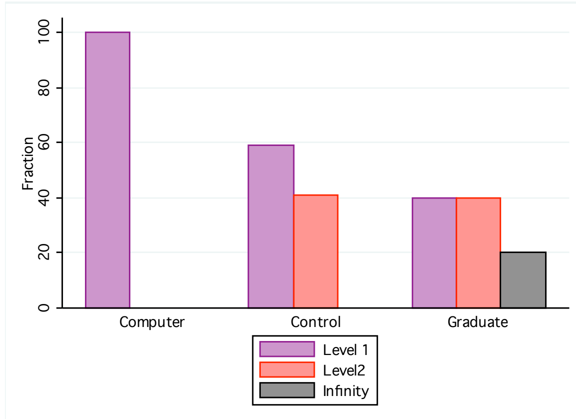


Figure 2: Distribution of levels in the Control, Graduate and Computer treatments, for those who are level 1 and above according to Nagel (1995).

Only individuals who have high objective cognitive levels can be expected to react to a change in the environment. Figure 2 shows the distribution of levels in all three treatments for those individuals who are levels 1 and above according to Nagel's classification. A Kolmogorov-Smirnov test rejects that the distribution of choices of those capable of at least one step of reasoning in the Baseline treatment is equal to the distribution of such choices in the Graduate treatment ($p = 0.032$). Further, both those distributions are different from the (degenerate) distribution in the Computer treatment where all subjects capable of one step of reasoning appear as level 1 players (a Kolmogorov-Smirnov test rejects the equality of distributions with $p < 0.001$ for both cases).[18]

Our findings above suggest that the observed distribution of levels can be manipulated by varying the expectations about other players's cognitive levels.

The Graduate and Computer treatments showed that the observed distribution of cognitive levels can be shifted up or down relative to the Control treatment. Both these treatments employed a between subject design. In the last treatment (Combo treatment) we use a within subject design, which allows us to identify and track the choices of those subjects who are

[18]This can be seen more dramatically in the Combo treatment where we use a within subject design and are able to follow individual subjects with high cognitive levels across treatment manipulations, with results significant at the 1% level (see Result 4).

level 1 or higher.

Result 2b: In the Combo treatment, the average choice of subjects decreases as the number of graduate students increase.

As the proportion of graduate students relative to that of computers in the group increases, the population our undergraduates are playing against is increasing in sophistication. If our undergraduate subjects believe that, then their choices should decrease. This is precisely what we observe at the aggregate level, as can be seen in Figure 3. A linear regression with average choice as the dependent variable and number of graduate students as the independent one shows that replacing a computer by a graduate student reduces the average choice by about 1 unit.[19]

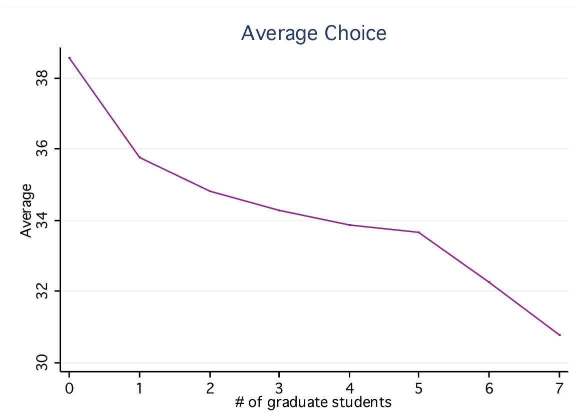


Figure 3: Average Choice by Number of Graduate Students in the Combo Treatment.

To further analyze the data and see how subjects respond to changes in the population they face, we use the cognitive hierarchy model of Camerer, Ho and Chong (2004). According to this model, levels of thinking are assumed to be distributed according to a Poisson distribution with parameter τ.[20] Following Camerer, Ho and Chong (2004), we report the best fitting estimate of τ of the data as well as the 90% confidence interval for τ from a

[19]The coefficient is $-.904$, the p-value is smaller than 0.001, while the R^2 is 91%.

[20]The process begins with level zero players, who are assumed to play according to a uniform distribution. Level k thinkers assume that the other players are distributed according a normalized Poisson distribution from level zero to level $k-1$. Hence they correctly predict the relative frequencies of levels zero through $k-1$, but may incorrectly believe that they are the only player of level k and that there are no players more sophisticated than they are. The estimation involves finding the value of that minimizes the difference between the observed sample mean and the mean implied by τ.

randomized resampling (with replacement) bootstrap procedure. In Table 5 we present our results in the Combo treatment for each configuration of computers and graduate students.

	Average	τ	Bootstrap 90% confidence interval
Combo treatment (0c,7g)	30.8	1.62	$[1.00, 2.25]$
Combo treatment (1c,6g)	32.3	1.44	$[1.07, 2.12]$
Combo treatment (2c,5g)	33.6	1.30	$[.79, 1.69]$
Combo treatment (3c,4g)	33.8	1.28	$[.99, 1.74]$
Combo treatment (4c,3g)	34.3	1.22	$[.73, 1.56]$
Combo treatment (5c,2g)	34.8	1.16	$[.78, 1.43]$
Combo treatment (6c,1g)	35.8	1.08	$[.98, 1.61]$
Combo treatment (7c,0g)	38.6	.81	$[.57, 1.42]$

Table 5: Estimation of tau in the Combo treatment using model of Camerer, Ho and Chong (2004).

In the cognitive hierarchy model, τ is a decreasing function of the average. Given that the average choices decrease as the number of graduate students increase, the τs will go in the opposite direction, as displayed in Table 5. The more interesting result is that the 90% confidence intervals for τ are moving windows with values generally increasing in the degree of sophistication of the group the undergraduates are playing against.[21] This provides support for the claim that the decrease in the average choice is not due to outliers, but rather reflects the choices of a significant part of the population.

These results indicate that on average subjects respond to changes in the composition of the population they face. The Combo treatment allows us to track individual behavior and look at how these aggregate results are generated.[22] Result 4 summarizes our findings.

Result 3: In responding to changes in the composition of groups in the Combo treatment, subjects can be classified into two types:

- **"Decreasing" types**: those who weakly decrease their choices as the proportion of

[21] We are using 90% confidence following Camerer, Ho and Chong (2004).

[22] While one may worry that the order of questions in the Combo treatment may indicate subjects that their responses should be monotonic, the data does not support this concern. Indeed, there are 45% of subjects whose choices were not monotonic (neither increasing nor decreasing) and 9% of subjects whose choices did not change as a response to the change in the composition of the group they were playing against.

graduate students in their group goes up, with at least one strictly decreasing choice.

- **"Random" types**: those who provide non-monotonic responses.

Table 6 presents some summary statistics for these two Random and Decreasing types, which represents 44.4% and 42.6% of the subjects, respectively. In addition, there are about 3.5% of Increasing types (defined as those who make weakly increasing choices with at least one strictly increasing choice) and 9.3% of Constant types who kept the same number (on average 31.6) for each configuration.

		Configuration (# computers, # graduate students)							
		(7c,0g)	(6c,1g)	(5c,2g)	(4c,3g)	(3c,4g)	(2c,5g)	(1c,6g)	(0c,7g)
Random (44.4%)	mean	46.4	42.9	42.6	43.6	45.3	46.8	45.9	45.4
	median	47	39.5	41.5	41	46.5	50.5	44	43.5
	st. dev.	22.6	16.5	16.2	18.7	16.9	18.8	21.8	22.6
Decreasing (42.6%)	mean	34.3	31.3	29.1	26.8	23.8	21.7	19.2	16.2
	median	33	30	29	25	24	21	18	15
	st. dev.	3.21	4.11	4.7	5.7	5.6	6.2	7.3	8.1

Table 6: Summary statistics of choices by type and configuration.

Table 6 shows that even in the $(7c, 0g)$ configuration, in which subjects are faced with 7 computers choosing numbers randomly, Random types fail to choose numbers that would correspond to level 1 behavior. This suggests that Random types are not capable of even one step of reasoning, which puts them into the category of non-strategic players (level zero).[23] Interestingly, in line with the standard assumption on level 0 play, the Random types average close to 50, as is the case in Agranov, Caplin and Tergiman (2010). On the contrary, the Decreasing people respond strategically in the $(7c, 0g)$ configuration with median and average choices of 33 and 34.3, respectively.[24] Moreover, as they are faced

[23]In the Combo treatment, about 7.4% of our subjects fit Nagel's definition of Level 0 players (this is similar to the proportions in the Graduate, Computer and Control treatments). Thus, Nagel's definition, by assuming that level 0 players play between 45 and 50 does not capture those who play randomly over the support. Our results suggest that the latter type of level 0 represent a larger fraction of those players.

[24]About 82.6% of the subjects in the Decreasing group choose numbers between 31 and 34 in the $(7c, 0g)$ configuration, which indicates that these people are at least level 1.

with more and more sophisticated population, the mean and the median of the numbers they choose decrease monotonically. Beyond the monotonicity of answers, there are other differences in the behavior of the subjects in the Random and Decreasing group. Subjects in the Random group also make larger changes between their answers for different configurations with a mean of 16 and a median of 10 compared with the subjects in the Decreasing group, who average a 2.6 units difference and a median change of 2 units. So, the Random types are not hovering around a fixed number but rather are changing their choices by large amounts. Further, over 75% of the subjects in the Random group change the direction of their answers at least 6 times over the 8 answers that they give in the game.[25]

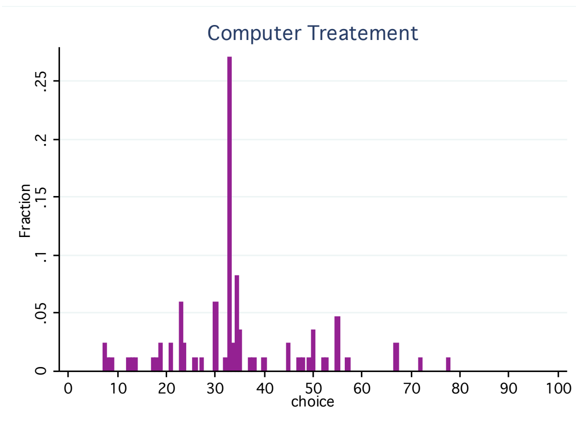
As Table 6 indicates, replacing a computer by a graduate student lowers the choice our undergraduates in the Decreasing group make by about 2.6 units. This implies that they believe graduate students are on average guessing 14.8. The close-to-linearity in the decrease of average choices (as the number of graduate students in the group they are playing against goes up) is an indication that subjects' beliefs about the choices of the graduate students is not dependent on how many of them they are playing against. In other words, the undergraduate students who are part of the Decreasing group believe that graduate students choose 14.8 on average, and this is true whether they are playing $(5c, 2g)$, $(1c, 6g)$ or any other configuraion.

Signrank tests confirm what is visually apparent from Table 6: the distribution of answers for all the pairwise different configurations are no different for the Random types ($p > 0.1$), while they are for the Decreasing types ($p < 0.001$). Consistent with Result 3, our findings from the Combo treatment indicates almost 50% of subjects are capable of at least one step of reasoning.

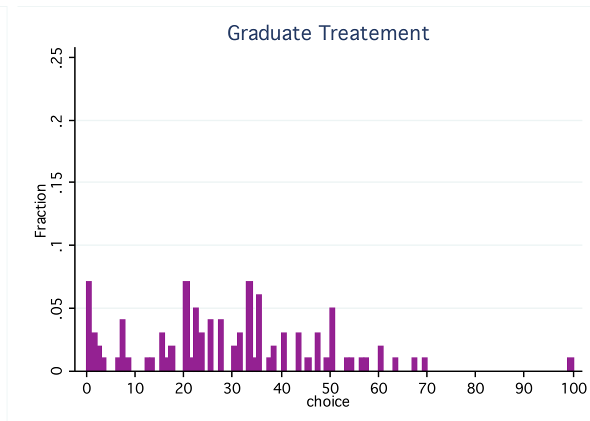
<u>Result 4:</u> **The strategy method used in the Combo treatment does not interfere with the strategic thinking of subjects.**

[25]The fact that this behavior is representative of close to half of the population may help explain why only little support has been found when experiments are designed to assess the stability of the level-k model across games, see Georganas, Heally and Weber (2011).

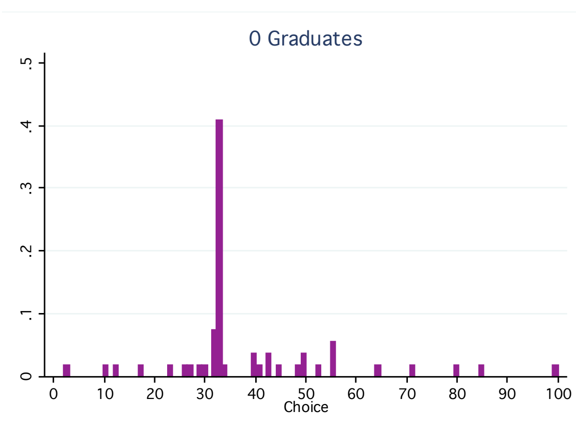
One question that may come to mind is whether the use of the strategy method in our Combo treatment had an effect on the behavior of our subjects. Since the design of the Combo treatment contains the Computer and Graduate treatments as a subcases, we answer this question by comparing the choices made in the Computer and Graduate treatments with that of the (7c, 0g) and (0c, 7g) configurations in the Combo Treatment. Figure 3 presents the histogram of the choices in the Computer and Graduate treatments as well as the distribution of choices in the Combo treatment with zero and seven graduate students, respectively. Table 7 presents the summary statistics of the choices observed in these treatments.



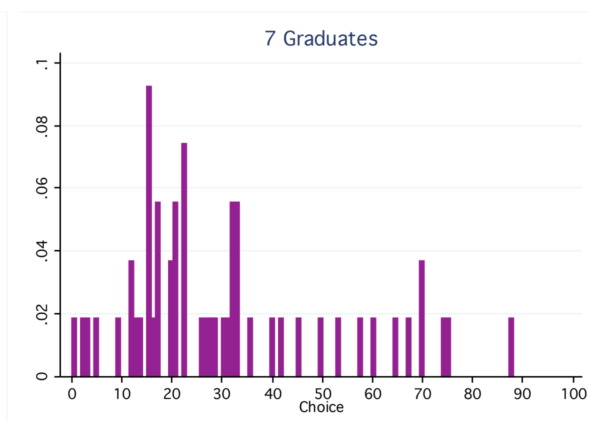
(a) Histograms of choices in the computer treatment.



(b) Histograms of choices in the graduate treatments



(c) Histograms of choices in the 0 graduates combo treatment.



(d) Histograms of choices in the 7 graduates combo treatment.

Figure 4: Histograms of choices in the Control, Graduate and Combo 0 and 7 graduates treatments.

As can be seen from Figure 3 and Table 7, the choices of our subjects in the (7c, 0g) configuration in the Combo treatment were close to those in the Computer treatment. A

	Mean Choice	Median Choice	Std. Dev.
Computer treatment	34.3	33	14.1
(7,0) configuration in Combo treatment	38.6	33	17.2
Graduate treatment	28.6	27	18.9
(0,7) configuration in Combo treatment	30.8	24	21

Table 7: Summary statistics of the Computer, Graduate and Combo treatments with 7 computers and 7 graduates students.

Ranksum test cannot reject the hypothesis that the distributions of choices are the same population ($p = 0.30$).[26] Similarly, the distribution of choices in the Graduate treatment is similar to the one in the $(0c, 7g)$ configuration in the Combo treatment ($p = 0.98$ in the Ranksum test).[27] [28]

To summarize, the strategy method used in the Combo treatment to elicit subjects' choices for various configurations of the group they are playing with did not alter the behavior in any significant way.

5 Conclusions

This paper has attempted to make one simple point that concerns the measurement of cognitive levels for subjects in games like the two-thirds guessing game. What we have shown is that the cognitive level chosen by subjects in these games is influenced not only by their ability to think strategically but also by their beliefs about the abilities of their cohorts. By manipulating these beliefs we have demonstrated that almost half of the subjects change their choices in the expected direction. The other half behave non-strategically. These results

[26] The p-value from a Kolmogorov-Smirnov test is 0.36.

[27] Note that since the Combo treatment was run at UBC while the Graduate and Computer treatments were run at NYU, the fact that there is no difference between the (0c,7g) and the Graduate treatment on one hand or, between the (7c,0g) and the Computer treatment on the other, also indicates that there is no difference in the subject populations in the two universities.

[28] To ensure that the rejection of the null isn't due to sample size differences, we conducted the following exercise: we cloned the data from the Combo treatment and obtained a sample double the original size (108). A Ranksum test shows that we still cannot reject the hypothesis that the data from the Combo treatment with zero and 7 graduate students are no different from the data from the Computer (85 observations) and Graduate (99 observations) treatments (with p-values well above 0.1 in both cases). The Kolmogorov-Smirnov test leads to the same conclusion.

show that for a large fraction of the population, cognitive type is endogenously determined and is a function of a player's belief about his opponents.

References

[1] Agranov, Marina, Caplin, Andrew and Chloe Tergiman. 2010. "The Process of Choice in Guessing Games." *Working Paper*.

[2] Arad, Ayala. 2011. "The Tennis Coach Problem: A Game-Theoretic and Experimental Study." *The B.E. Journal of Theoretical Economics*, forthcoming.

[3] Bosch-Domenech, Antoni, Montalvo, Jose, Nagel, Rosemarie and Albert Satorra. 2002. "One, Two, (Three), Infinity, ... : Newspaper and Lab Beauty-Contest Experiments." *Working Paper.*

[4] Buhren, Christoph and Frank Bjorn. 2010. "Chess players' performance beyond 64 squares: A case study on the limitations of cognitive abilities transfer." *Working Paper.*

[5] Burchardi, Konrad, and Stefan Penczynski. 2010. "Out of Your Mind: Eliciting Individual Reasoning in One Shot Games." *Working Paper*.

[6] Camerer, Colin, Ho, Teck-Hua, and Juin-Kuan Chong. 2004. "A Cognitive Hierarchy Model of Games." *The Quarterly Journal of Economics*, 119(3): 861-898.

[7] Choi, Syngjoo. 2011. "A Cognitive Hierarchy Model of Learning in Networks." *Working Paper.*

[8] Coricelli, Giorgio, and Rosemarie Nagel. 2009. "Neural correlates of depth of strategic reasoning in medial prefrontal cortex." *Proceedings of the National Academy of Sciences: Economic Sciences*, 106(23): 9163-9168.

[9] Costa-Gomes, Miguel, and Vincent Crawford. 2006. "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study." *The American Economic Review*, 96(5): 1737-1768.

[10] Crawford, Vincent, Costa-Gomes, Miguel and Nagore Iriberri. 2010. "Strategic Thinking." *Working Paper.*

[11] Georganas, Sotiris, Healy, Paul and Roberto Weber. 2011 "On the Persistence of Strategic Sophistication". *Working Paper.*

[12] Grosskopf, Brit and Rosemarie Nagel. 2008. "The Two-Person Beauty Contest." *Games and Economic Behavior* 62: 93–99.

[13] Ho, Teck-Hua, Camerer, Colin, and Keith Weigelt. 1998. "Iterated Dominance and Iterated Best-response in p-Beauty Contests." *The American Economic Review*, 88: 947-969.

[14] Nagel, Rosemarie. 1995. "Unraveling in Guessing Games: An Experimental Study." *The American Economic Review*, 85(5): 1313-1326.

[15] Palacios-Huerta, Ignacio and Oscar Volij. 2011. "Field Centipedes." *The American Economic Review*, forthcoming.

[16] Kocher, Martin, Sutter, Mattias and Florian Wakolbinger. 2009. "Social learning in beauty-contest games." *Working Paper.*

[17] Stahl, Dale, and Paul Wilson. 1995. "On Players Models of Other Players: Theory and Experimental Evidence." *Games and Economic Behavior*, 10(1): 218-254.

[18] Strzalecki, Tomasz. 2010. "Depth of Reasoning and Higher Order Beliefs." *Working Paper.*

[19] Weber, Roberto. 2003. "Learning with no feedback in a competitive guessing game." *Games and Economic Behavior*, 44(1): 134-144.

6 Appendix A: Instructions for the Combo Treatment

Recently, a group of 8 graduate Ph.D. students from the department of Economics at NYU with training in these types of games played the following game. Each graduate student in the group chose a number between 0 and 100. That student who was the closest to the 2/3 of the average of all the chosen numbers (including their own) won $10.

Your task

Choose a number between 0 and 100. You will win $10 if your chosen number is closest to 2/3 times the average of all the chosen numbers of the players in your group.

Your Group

There are 8 members in your group: you plus seven others. The seven other players will be a mixture of computer players and those graduate students mentioned above.

- *Computers.* Each of the computers will choose a random number between 0 and 100 with each number being equally likely and with each computer choosing independently from the others.
- *The graduate students.* We will randomly select a certain number of responses from the set of 8 graduate student responses we have (from when they previously played the game). It is important to note that the responses we randomly choose are from those choices the graduate students made when they knew they were playing against each other. They never played against computers and will not be engaged in your decision problem.

Each of the rows on the sheet below describes the number of computers and graduate students in your group.

You will now have to choose a number between 0 and 100 for each of these rows. In other words, you will need to choose a number between 0 and 100 in eight different circumstances where the fraction of computer and graduate student choices you are playing against varies from 0 computer choices and 7 graduate student choices to 7 graduate student choices and 0 computer choices.

Notice that you are not playing against other people in the room. Each of you is playing against your group, which consists of you, a certain number of computers and a certain number of graduate students. So, any number of you may win $10. It all depends on your choice and the choices of the computers and graduate students in the relevant computer/graduate student group you are in.

To determine your payment, once we collect all the sheets, we will randomly select one of those rows (using a dice) to "count." It will be your answer in that row that will matter. You will win $10 if your chosen number is closest to 2/3 times the average of all the chosen numbers of the players in your group.

Your Group	**Your Number**
You, 7 computers, 0 graduate students	
You, 6 computers, 1 graduate students	
You, 5 computers, 2 graduate students	
You, 4 computers, 3 graduate students	
You, 3 computers, 4 graduate students	
You, 2 computers, 5 graduate students	
You, 1 computers, 6 graduate students	
You, 0 computers, 7 graduate students	