Beliefs and Endogenous Cognitive Levels: an Experimental Study*

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

February 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 2/3 guessing game, individuals' choices do not always reflect their full cognitive abilities, but rather are a reflection of their full cognitive abilities and their beliefs on others' abilities. Hence, a subject's true cognitive level may be greater 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.

JEL Codes: C72, C92, D83.

Keywords: Guessing Game, Beliefs, Level-k theory.

[†]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 cognitive level as endogenously determined by his expectations about the (also endogenously determined) cognitive levels of others.[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 2/3 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 2/3 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.

We will refer to the last element as an agent's "objective cognitive level." In other words, the objective cognitive level is an upper bound on what someone is capable of acting like. If

---

[1]Theoretically, the only model that we are aware of that has endogenous choice of cognitive levels is Choi (2006) 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.

one could look inside each person's brain one could, in principal, measure how many levels each person is capable of and construct this objective level distribution. However, when the game is actually played, people may choose to behave as a lower level player if they believe that others in their group are distributed in a manner different from the objective distribution, for example if the objective distribution stochastically dominates their subjective guess of it. Therefore, the observed level distribution might differ from the objective level distribution.[2]

The goal of this paper is to illustrate the difference between the objective and the observed level distributions using a number of manipulations regarding the expectations about the cognitive levels of other players. We start by having undergraduate students play the 2/3 guessing game against a group of graduate students who have training in these sorts of games ("Graduate" treatment). In this treatment, if undergraduates believe that trained graduate students' levels of cognition in this task are higher than the average of their peers, they should exhibit higher cognitive levels when playing against the former than when playing against the latter. This is precisely what we observe: relative to the "Control" treatment, in which undergraduate students are playing against each other, we observe a significant shift in the distribution of cognitive levels towards more sophisticated behavior when undergraduate students are playing against graduate students. That is, the distribution of choices in the Graduate treatment is closer to the objective level distribution compared with the distribution of choices in the Control treatment. 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.

Our main contribution is that we control beliefs and highlight the role that they play in the 2/3 guessing game. We show that subjects may behave differently depending on their

---

[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.

beliefs regarding the sophistication of other players. In other words, we demonstrate the difference between the observed and objective cognitive level of subjects by manipulating (and thus controlling) beliefs about their opponents. Since experience is held fixed, the difference in behavior is entirely attributable to the differences in beliefs. Overall, our results show that in at least some strategic situations, subjects do not necessarily function at their highest cognitive level but rather are poised to if they believe that others will do so. This behavior describes almost half of the population.[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 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.

---

[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 2/3 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.

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 2/3 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 ((7c, 0g) configuration) as in the Computer treatment, to 0 computers and 7 graduate students ((0c, 7g) configuration), as in the Graduate treatment. We start by showing that the choices from the (7c, 0g) configuration in the Combo treatment match the choices from our Computer treatment. There is also no statistical difference between the answers from the (0c, 7g) 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 2/3 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 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 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 2/3 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 2/3 guessing game. The authors use a two-person 2/3 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 2/3 guessing game vary either with repeated play or with subject pools. For example, Weber (2003) has subjects play a series of 2/3 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.

The rest of the paper is organized as follows. In Section 2 we describe the experimental procedure and design. In Section 3 we analyze the experimental results. The conclusions are in section 4.

2 Experimental Design

The experiments were conducted in classrooms of undergraduate students at New York University and at the University of British Columbia.[4] In total, 359 students participated in 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 2/3 times the average of all chosen

---

[4]The UBC students played in the Combo treatment. The NYU students played the Control, Graduate and Computer treatments. See footnote 18 in Result 5 of Section 3 for evidence that the behavior of NYU and UBC students did not significantly differ.

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 provide participants with enough incentives to think as hard as they can about the game being played and behave closer to their objective cognitive levels. To achieve this we first conducted a 2/3 guessing game with a group of 8 graduate students from the department of economics at NYU using the instructions from our Control treatment. We then conducted a 2/3 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 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 two thirds 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 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.[5] 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.

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. 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.[6] 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.[7] This treatment allows

---

[5]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.

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

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

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.[8]

3   Results

Result 1: 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)[9] and compare this distribution to Nagel's data.

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.

---

[8]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.

[9]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 13 and 16.

	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).

The main hypothesis of this paper is that the observed distribution of cognitive levels is endogenous and can be manipulated by varying the expectations about other players's cognitive levels. To test this hypothesis, we compare the choices of our subjects across the Control, Computer and Graduate Student treatments. Our findings are presented in Result 2 below.

Result 2: 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. In addition, the proportion of subjects capable of at least one level of reasoning is about one half.
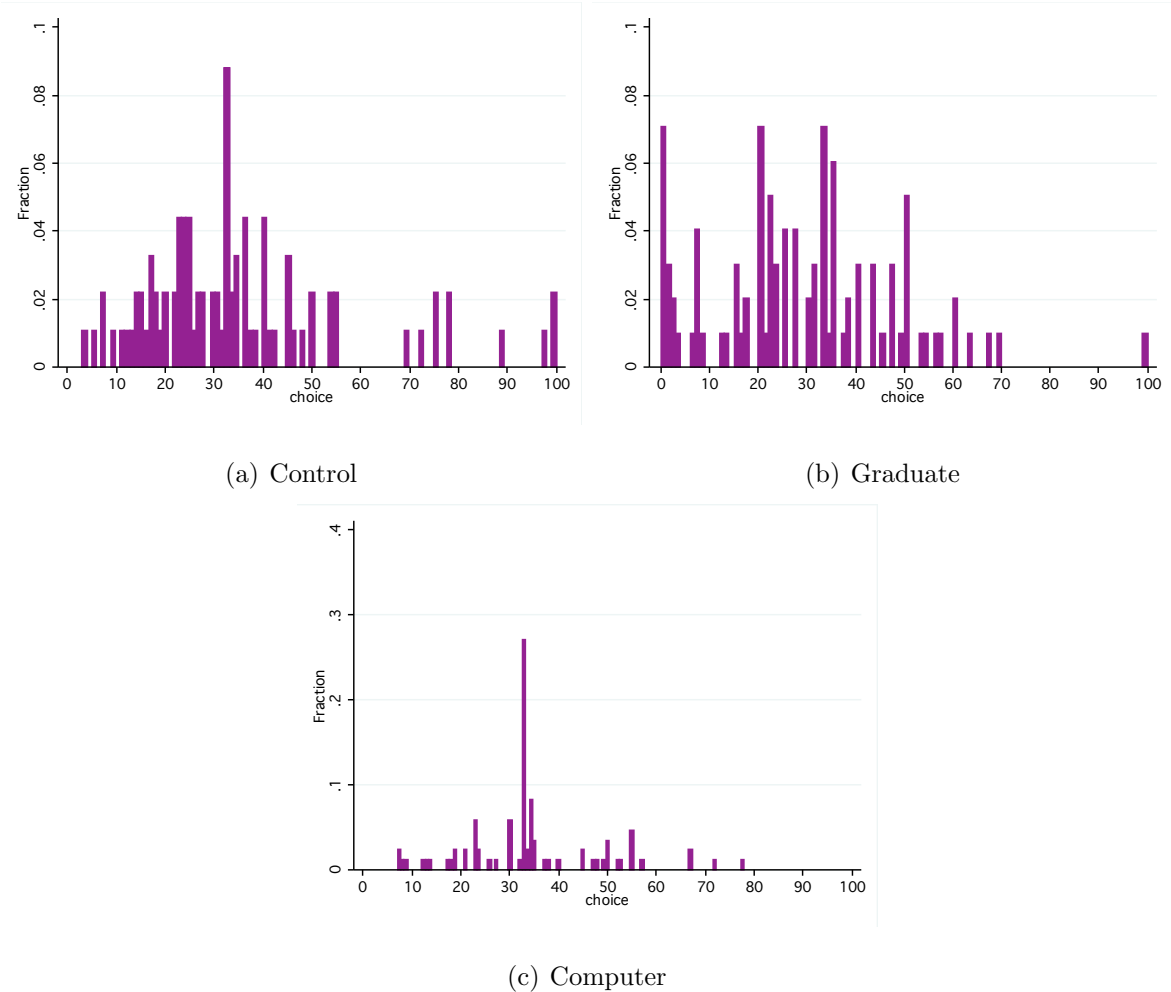

(a) Control

(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 Control, Graduate and Computer treatments. In Tables 3 and 4 we present summary statistics of the choices observed in these three treatments, and the distribution of cognitive levels in these treatments according to the Nagel (1995) classification, respectively.[10]

	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%	
Level 2	18%	20%	
Level 3	8%	5%	49%
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.

Table 4 presents some interesting results. First, note that in the Computer treatment any subject with an objective cognitive level of at least 1 should have an observed cognitive level of exactly 1 and should choose (about) 32, since 32 is the best response to the computers

---

[10]Since all higher level reasoners should act as level 1 players in the Computer treatment, we cannot separately identify them in Table 4: together they represent 49% of the population.

choosing randomly. If we interpret Nagel's classification strictly, however, many of those choosing below that number would be classified as higher levels. This is actually a false classification in our context since their choices indicate confusion and not sophistication: any truly strategic subject would know that his best response to a set of computers is 32 and not something lower. For the other treatments, to find the number of subjects who are level 1 or higher we must sum all of those classified as at least level 1 since by not controlling their beliefs, lower choices can be rationalized. In the Computer, Graduate and Control treatments the number of subjects classified as being level 1 or more is 49%, 55% and 51%, respectively. A test of proportions indicates that these percentages are not different (all pairwise p-values are greater than 10%).
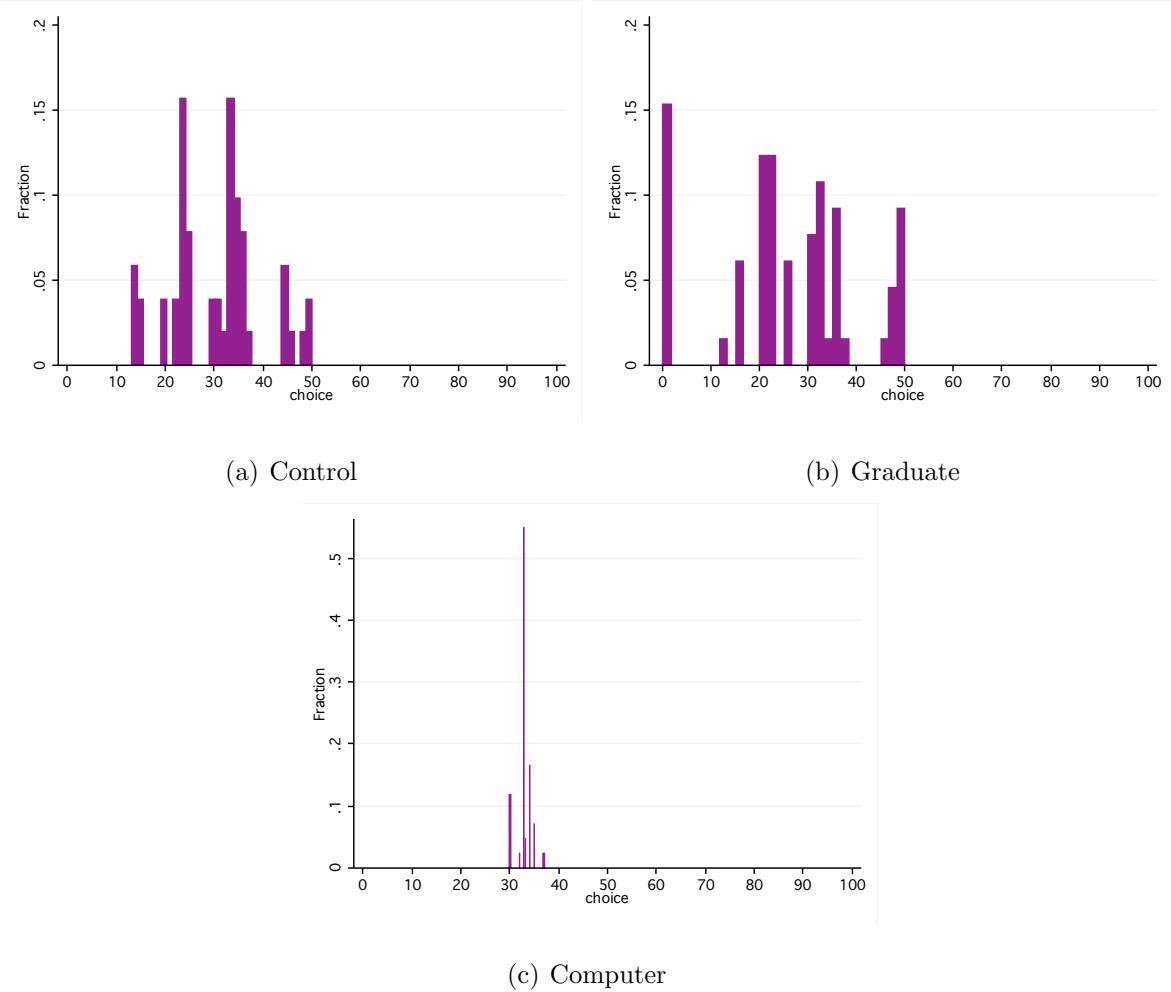

(a) Control

(b) Graduate

(c) Computer

Figure 2: Histograms of choices in the Control, Graduate and Computer treatments, for those who are level 1 and above.

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

To demonstrate this first note that only individuals who have high objective cognitive levels (level 1 or above) can be expected to react to a change in our treatments. Those who are level 0 (or unclassified) either do not respond to the sophistication of their opponents or do so in unpredictable ways. Hence, the proper test of whether subject behavior changes as we change treatments is to compare the distribution of choices of those subjects who are classified as being of level 1 or above. This is done in Figure 2 which shows the distribution of choices in all three treatments for those individuals who are levels 1 and above according to Nagel's classification. A Kolmogorov-Smirnov test rejects the hypothesis that the distribution of choices of those capable of at least one step of reasoning in the Control 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). This same result reappears even 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).

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 level 1 or higher.

Result 3: 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.[11]
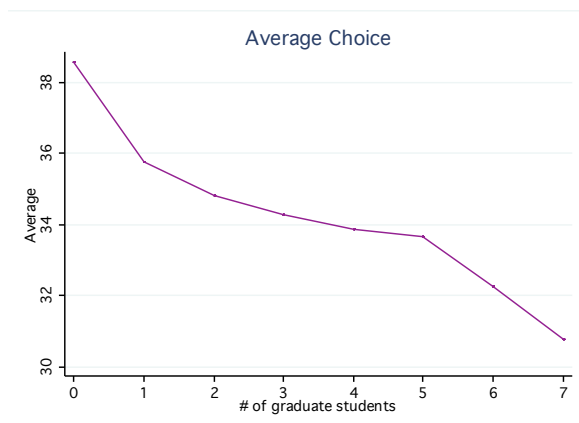
Figure 3: Average Choice by Number of Graduate Students in the Combo Treatment.

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.[12] 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

---

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

[12]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

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.[13] 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. Result 4 summarizes our findings.

---

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.

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

Result 4: In responding to changes in the composition of groups, subjects can be classified into two types:

- **"Decreasing" types**: those who weakly decrease their choices as the proportion of 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)
	mean	46.4	42.9	42.6	43.6	45.3	46.8	45.9	45.4
Random (44.4%)	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
	mean	34.3	31.3	29.1	26.8	23.8	21.7	19.2	16.2
Decreasing (42.6%)	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).[14] 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.[15] Moreover, as they are faced 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.[16]

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

---

[14]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.

[15]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.

[16]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).

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.
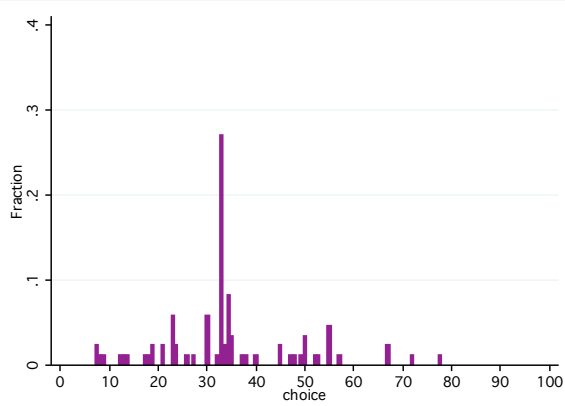
Result 5: The strategy method used in the Combo treatment does not interfere with the strategic thinking of subjects.

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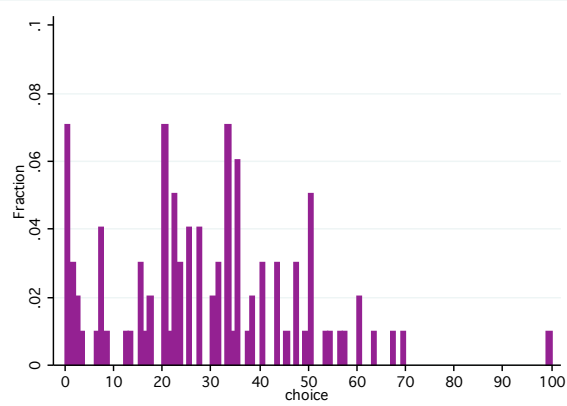
	Mean Choice	Median Choice	Std. Dev.
Computer treatment	34.3	33	14.1
(7,0) configuration in Combo treatment	38.6	33	17.2
Graduate treatment	28.6	27	18.9
(0,7) configuration in Combo treatment	30.8	24	21

Table 7: Summary statistics of the Computer, Graduate and Combo treatments with 7 computers and 7 graduates students.
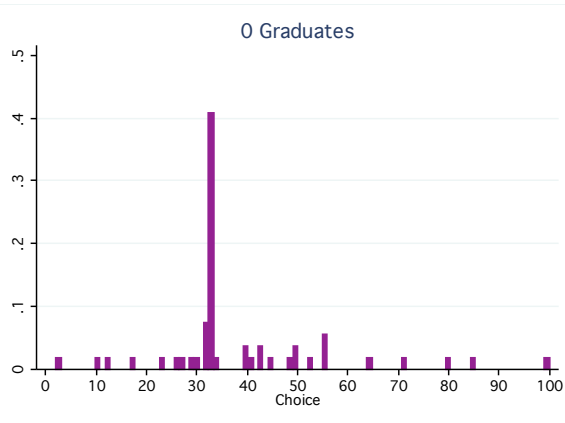
As can be seen from Figure 3 and Table 7, the choices of our subjects in the $(7c, 0g)$
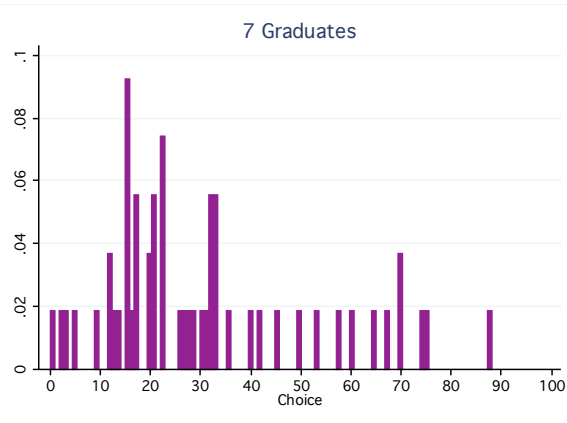
(a) Histograms of choices in the computer treat- (b) Histograms of choices in the graduate treat-
ment.   ments

(c) Histograms of choices in the 0 graduates combo (d) Histograms of choices in the 7 graduates combo
treatment.   treatment.

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

configuration in the Combo treatment were close to those in the Computer treatment. A
Ranksum test cannot reject the hypothesis that the distributions of choices are the same
population ($p = 0.30$).[17] 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).[18]

---

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

[18]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

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.

4 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 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.

---

difference in the subject populations in the two universities.

References

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

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

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

[4] 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.

[5] Choi, Syngjoo. 2006. "A Cognitive Hierarchy Model of Learning in Networks." *Working Paper.*

[6] 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.

[7] 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.

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

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

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

[11] 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.

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

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

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

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

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

5 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	