

# Level- $k$ Reasoning in a Generalized Beauty Contest\*

Dmitry Shapiro Xianwen Shi Artie Zillante

August 31, 2011

## Abstract

We study how the predictive power of level- $k$  models changes as we perturb the classical beauty contest setting along two dimensions: the strength of the coordination motive and the information symmetry. We use a variation of the Morris and Shin (2002) model as the unified framework for our study, and find that the predictive power of level- $k$  models varies considerably along these two dimensions. Level- $k$  models are successful in predicting subject behavior in settings with symmetric information and a strong coordination motive. However, the predictive power of level- $k$  models is significantly weakened when private information is introduced or the importance of the coordination motive is decreased.

## 1 Introduction

The experimental literature on beauty contests and related guessing games has documented substantial evidence that individuals tend to have a limited degree of strategic sophistication, especially in settings where the strategic reasoning is not straightforward. This is best illustrated by the “ $p$ -beauty contest” in which participants choose a number between 0 and 100 and whoever picks the number closest to a multiple  $p$  of the group average wins a prize. When  $p$  is less than one the game can be solved by iterative elimination of strictly dominated strategies, and the unique equilibrium is where every player chooses 0. In order to reach this equilibrium subjects need to go through a large number of rounds of elimination of dominated strategies. The experimental literature on beauty contests, however, shows that subjects usually perform one to three rounds of elimination and that their behavior is consistently different from the equilibrium prediction.

The theory of level- $k$  reasoning, first proposed by Stahl and Wilson (1995) and Nagel (1995) with further extensions by Ho, Camerer, and Weigelt (1998), Costa-Gomes, Crawford and Broseta

---

\*Shapiro and Zillante: Belk College of Business, University of North Carolina, 9201 University City Boulevard, Charlotte NC 28223-0001, USA; email address: dashapir@uncc.edu, azillant@uncc.edu. Shi: Department of Economics, University of Toronto, 150 St. George Street, Toronto, Ontario M5S 3G7, Canada; email address: xianwen.shi@utoronto.ca. We would like to thank Stephen Morris for many helpful comments and constant feedback. We are also grateful to Vincent Crawford, Kirill Evdokimov, and Pei-Yu Lo for helpful comments and suggestions. This research was supported by funds provided by a Childress Klein Research Grant, the Belk College of Business, the Connaught start-up fund at University of Toronto and a SIG award from SSHRC of Canada.

(2001) and Costa-Gomes and Crawford (2006), can be used to rationalize subject behavior in the  $p$ -beauty contest. The level- $k$  model is based on the presumption that subjects' behavior can be classified into different levels of reasoning. The zero level of reasoning,  $L0$ , corresponds to non-strategic behavior when strategies are selected at random without forming any beliefs about opponents' behavior. In the literature  $L0$  is typically considered to be a person's model of others rather than an actual person. Level-1 players,  $L1$ , believe that all their opponents are  $L0$  and play a best response to this belief. Level-2 players,  $L2$ , play the best response to the belief that all their opponents are  $L1$  and so on. For example, when  $p$  is equal to  $2/3$  in the beauty contest, level-1 players choose 33 and level-2 players choose 22. As was shown in Nagel (1995) and many other papers, there is indeed a salient pattern of levels of reasoning in the beauty contest setting.

While level- $k$  thinking is not particularly unique to the beauty contest (see e.g. Costa-Gomes and Crawford, 2006), the structure of the game and its simplicity are very conducive to this type of behavior. Success in the beauty contest largely depends on a person's ability to correctly predict the average choice made by others. This explicitly forces individuals to think about decisions of other players. Moreover, the symmetry of information makes this task relatively simple, which can further encourage participants to focus on the behavior of others.

In many real applications, however, market participants often have access to both public and private information on the underlying fundamentals, and choose actions that are not only responsive to peer action choices but also appropriate to the fundamentals. A natural question then arises: how will level- $k$  models perform beyond the classical beauty contest setting?

To answer this question, we introduce a framework which generalizes the classical beauty contest setting along two dimensions. First, it allows players to have private information that is relevant for their action choice. Second, it allows the importance of coordination to change so that the ability of correctly guessing other players' actions may have a different impact on players' payoffs. We then analyze how the predictive power of the level- $k$  models varies along these two dimensions.

The generalized framework that we use for our study is a modification of the Morris and Shin (2002) model (hereafter MS) on the social value of public information. In our setting, just as in Morris and Shin, the agents' payoff is determined by two criteria: how well an agent's action matches an unknown state of the world and how well his action matches the average actions of other agents. The relative importance of both factors can be varied within the model. In particular, as the latter becomes more important it makes the coordination motive of the game stronger. Agents in our model receive two signals about the (unknown) underlying state. If both signals are public the information is symmetric. If one signal is public and the other is private (as in the original Morris and Shin setting) then the information is asymmetric and, in particular, different participants have different information.

Based on this framework we design several experimental treatments that differ from each other in the symmetry of information and in how important it is to predict the average action of other players. Our main findings are as follows. First, in aggregate we find that subjects place less weight on the public signal than the MS model predicts. We show that this is consistent with the theoretical prediction of level- $k$  models. An important implication is that, if agents have limited

cognitive ability, the detrimental effect of increased public disclosure on social welfare may not be as strong as the MS model predicts.

Second, we compare individual subjects' behavior with level- $k$  predictions. We find that in treatments with public information *and* a strong coordination motive subjects' behavior is consistent with level- $k$  reasoning. In these treatments the percentage of individuals playing according to level- $k$  models is as high as 85%, which is in sharp contrast with private information treatments or treatments with weak coordination motive where it never reaches 40%. Another individual level finding is that the most commonly used level of reasoning in our data is  $L1$ . This result holds robustly across most of the treatments. The most notable exception is the treatment with public signals and a strong coordination motive where the most common level was  $L2$ . Our last individual level result is that we do not find any evidence of level- $k$  reasoning being more prevalent in the beginning of the treatment. Thus our data do not support the common interpretation of level- $k$  behavior as a model to describe subjects' initial responses.

Finally, we perform maximum likelihood (ML) estimation of the level- $k$  model and a closely related cognitive hierarchy model (CH) model introduced in Camerer, Ho, and Chong (2004). We find that with few exceptions both models predict subjects behavior better than NE. The models are particularly successful in public information treatment with the strong coordination motive. In fact, regardless of the information structure when the coordination motive is stronger the estimated shares of strategic types (i.e. the types that are not level-0) are higher and are more likely to be significant. Comparing the standard level- $k$  model and CH we conclude that the CH performs better in matching the data. This is particularly surprising since the CH has fewer parameters (one for public information treatments and two for private information treatments) than the standard level- $k$  model.<sup>1</sup>

Overall, our analysis highlights the strengths and limitations of level- $k$  models. The modified Morris and Shin framework used in our study is considerably more complicated than those typically used in the level- $k$  literature. Despite this complexity level- $k$  models are very successful in predicting subjects' behavior in settings that are close to the classical beauty contest, such as when the coordination motive is strong and information is symmetric. At the same time we find that the predictive power of level- $k$  models diminishes as we move away from the classical setting and either weaken the coordination motive or introduce private information.

Our experimental findings also have important policy implications. The key insight in the analysis of Morris and Shin (2002) is that in equilibrium players often place too much weight on the public signal relative to the weight that would be used by the social planner. Therefore, individual information aggregation is not socially efficient and enhanced public disclosure could hurt social welfare. However, our theoretical analysis of level- $k$  reasoning shows that limited cognitive ability, either due to limited level of reasoning or incapability of Bayesian updating, necessarily leads to subjects underweighting the public signal compared to the equilibrium prediction. We find in our experiment that subjects indeed put less weight on the public signal than the theory

---

<sup>1</sup>Gneezy (2005) applies the framework of cognitive hierarchy to analyze first-price and second-price common value auctions with complete information, and finds evidence supporting the CH theory.

predicts. This finding is also documented independently in a recent experimental study by Cornand and Heinemann (2009). This implies that limited cognitive ability can limit the detrimental effect of increased public disclosure.

The rest of the paper is organized as follows. In Section 2 we discuss relevant literature on level- $k$  thinking and the MS model. In Section 3 we provide a theoretical background for our study which is largely based on the MS model. We derive the prediction of level- $k$  models in this setting and show that subjects with limited cognitive ability will put less weight on the public signal than the equilibrium predicts. Section 4 provides details of our experimental design and various treatments. Our experimental results are reported in Section 5. Section 6 then concludes and the experimental instructions are given in the Appendix.

## 2 Literature Review

Our experimental study contributes to the existing literature on the classical beauty contest beginning with Nagel (1995), who first documents the clear pattern of level- $k$  thinking in subject behavior. Ho, Camerer, and Weigelt (1998); Bosch-Domenech, Montalvo, Nagel, and Satorra (2002); Costa-Gomes and Crawford (2006); and Crawford and Iriberri (2007a,b), among others, have further developed and applied level- $k$  models to beauty contests and related settings. However, most of the existing literature focuses on games with complete information. Notable exceptions are Crawford and Iriberri (2007a), who applied level- $k$  reasoning to first- and second-price auctions, and a recent independent work by Cornand and Heinemann (2009), which is closely related to our paper.

Cornand and Heinemann (2009) conduct experiments within the framework of the MS model and find that subjects put less weight on the public signal than the theory predicts. By assuming that all subjects use a common level of reasoning, they find that subject behavior is consistent with the second level of reasoning ( $L2$ ). They further argue that, if all subjects behave according to  $L2$ , the welfare result in Morris and Shin does not hold: increasing the precision of public information is always beneficial. Their paper and ours share the same theoretical framework and both find that the increased disclosure of public information is less detrimental than the theory predicts. But there are important differences. They exclusively focus on the welfare implications of public disclosure, whereas our main focus is to test the performance of level- $k$  models across settings with different information and payoff structures. Moreover, we assume that the population consists of a mixture of different levels, whereas for the purposes of Cornand and Heinemann it was sufficient to assume a common level.

The framework underlying our experimental study is first developed by Morris and Shin (2002) to evaluate the value of public information on social welfare in a coordination environment. Subsequently, Angeletos and Pavan (2007) generalize their analysis of the social value of information by allowing both strategic complementarity and strategic substitutability among agents' actions. The Morris and Shin framework has been applied to many different settings including asset pricing (Allen, Morris and Shin, 2006, Bacchetta and Wincoop, 2005), venture capital (Angeletos, Lorenzoni and Pavan, 2007) and political science (Dewan and Myatt, 2007, 2008).

### 3 Theoretical Background

This section provides a theoretical background for our study. The primary goal of our paper is to analyze performance of level- $k$  reasoning in a setting which is similar to the classical beauty contest yet allows us to vary the importance of the coordination motive and information structure. For this purpose we use the Morris and Shin (2002) framework as a basis for our experimental analysis. However, since the original MS model cannot be directly implemented in the lab due to such assumptions as continuum of agents and improper uniform distribution we, first, need to modify it to adapt it to experimental environment. This is done in section 3.1. In section 3.2 we use the modified MS framework to derive predictions of the level- $k$  model.

#### 3.1 Modified Morris-Shin Model

There are  $n$  ex-ante identical agents,  $i = 1, \dots, n$ . Agent  $i$  chooses an action  $a_i \in \mathbb{R}$ . The payoff function for agent  $i$  is given by

$$u_i(a_i, \bar{a}_{-i}, \theta) = C - (1 - r)(a_i - \theta)^2 - r(a_i - \lambda \bar{a}_{-i})^2, \quad (1)$$

where  $C$  is a constant,  $\theta$  represents the underlying state,  $r$  and  $\lambda$  are constants between 0 and 1, and  $\bar{a}_{-i}$  is the average action of  $i$ 's opponents:  $\bar{a}_{-i} = \frac{1}{n-1} \sum_{j \neq i} a_j$ .

The payoff function has three terms. The first one is a constant  $C$  and is the highest payoff the individual can possibly receive. The second term reflects the loss from mismatching the underlying state  $\theta$  and is simply the square of the distance between  $\theta$  and  $a_i$ . The third term is the ‘‘beauty contest’’ term. It measures the loss from mismatching the average action of opponents  $\bar{a}_{-i}$  which is scaled by  $\lambda$ . The parameter  $r$  measures the relative importance of coordinating with opponents’ actions versus matching the underlying state. When  $\lambda = 1$  and  $C = 0$  the game becomes the coordination game specified in MS. When  $r = 1$  and  $\lambda < 1$  the game becomes similar to the beauty contest in the sense that subjects only need to match  $\lambda$  times the average of other players’ actions. Unlike the beauty contest, however, everyone, not just the player whose guess is the closest to the target, receives a non-negative payoff.

Our payoff function differs from the MS one in three ways. First, we consider a setting with a finite number of players while in MS there is a continuum of players. Second, we introduce the term  $\lambda$  inside the payoff function to match the classical  $p$ -beauty contest. Third, the payoff function in MS is always negative, which is difficult to implement in the laboratory. By adding a positive constant  $C$  to the original payoff function we allow participants’ payoffs to be positive without altering equilibrium predictions.

As in MS, before taking actions, agent  $i$  will receive two signals about  $\theta$  and we assume that both signals have the same precision  $\alpha$ . The first signal  $y$  is always public and is given by

$$y = \theta + \eta, \quad \eta \sim N(0, 1/\alpha). \quad (2)$$

As for the second signal,  $x_i$ , it can be either public or private. If it is private, then

$$x_i = \theta + \varepsilon_i, \quad \varepsilon_i \sim N(0, 1/\alpha), \quad (3)$$

and  $\eta$  and  $\varepsilon_i$  are independent. If it is public, then it is the same across agents and is given by

$$x_i = \theta + \varepsilon, \quad \varepsilon \sim N(0, 1/\alpha).$$

Again  $\eta$  and  $\varepsilon$  are independent. After receiving  $x_i$  and  $y$ , agent  $i$  chooses action  $a_i$ . Morris and Shin assume that  $\theta$  is distributed with the improper uniform distribution over the real line in which case the expected value of  $\theta$  given  $x_i$  and  $y$  is

$$\mathbb{E}_i(\theta|x_i, y) = \frac{y + x_i}{2}. \quad (4)$$

Following the same procedure as in MS we can show that when  $x_i$  is private the unique equilibrium is linear and is given by

$$a_i(y, x_i) = \frac{1-r}{2-\lambda r}x_i + \frac{1-r}{(2-\lambda r)(1-\lambda r)}y. \quad (5)$$

When signal  $x_i$  is public, the unique Nash equilibrium is

$$a_i(y, x_i) = \frac{1-r}{2-2\lambda r}x_i + \frac{1-r}{2-2\lambda r}y. \quad (6)$$

Notice, in particular, that when  $\lambda < 1$  and  $r = 1$  the NE is 0, as in the beauty contest.

A major difficulty of implementing the MS setup in the lab is to generate  $\theta$  according to the improper uniform distribution. To deal with this problem we adopted the following strategy. We generated  $\theta$  using the uniform distribution on interval  $[a, b]$  and then given  $\theta$  we generated the signals  $y$  and  $x_i$  according to (2) and (3). After that we normalized state  $\theta$  and signals  $(x_i, y)$  by subtracting  $y$  from each of them, so that  $\theta_n = \theta - y$ ,  $x_{i,n} = x_i - y$  and  $y_n = 0$ .

Since the prior of  $\theta$  has a bounded support, the formula (4) to obtain  $E(\theta|x_i, y)$  may not be valid, and thus the NE would no longer be given by (5) and (6). However, normalized signals are immune to this problem. By the definition of  $y$ , we have  $\theta_n = -\eta$  and  $x_{i,n} = \varepsilon_i - \eta$ . Since both  $-\eta$  and  $\varepsilon - \eta$  are normally distributed, by the standard formula for the conditional distribution of normally distributed random variables we have

$$\mathbb{E}(\theta_n|x_{n,i}) = \mathbb{E}(-\eta|\varepsilon_i - \eta) = \frac{1}{2}(\varepsilon_i - \eta) = \frac{x_{i,n}}{2}.$$

Given that  $y_n = 0$  this is the same as (4). Therefore, when agents observe normalized signals the MS logic and the equilibrium derivations remain valid. In the experimental design section we provide more details on how the normalization was implemented.

### 3.2 Calculating Levels of Reasoning

Within the setting introduced in the previous section we derive actions that correspond to different levels of reasoning. From now on we assume that signals and the state are normalized and with slight abuse of notation we will use  $\theta, x_i$  and  $y (= 0)$  to denote the normalized signals. It is convenient to introduce the variable  $\mu = 1/2$  so that player  $i$ 's updated estimate of the state can be written as  $\mathbb{E}_i[\theta] = \mu x_i$ .

Player  $i$  chooses  $a_i$  to maximize (1) and from the first-order condition the best response is

$$a_i^* = (1 - r) \mathbb{E}_i [\theta] + r\lambda \mathbb{E}_i [\bar{a}_{-i}].$$

Except for the non-strategic  $L0$  type, agents with different levels of reasoning will form different beliefs about  $\mathbb{E}_i [\bar{a}_{-i}]$  and will choose an action accordingly.

The first step in calculating  $Lk$  actions is to define the behavior of  $L0$ . In the literature type  $L0$  is usually viewed as the starting point of a player's analysis of others' actions, so it should be unsophisticated and non-strategic (see e.g. Crawford and Iriberri, 2007). In our paper we assume that  $L0$ 's actions are uniformly distributed between the two signals. Under this assumption  $L0$ 's behavior is indeed unsophisticated and serves as a natural focal point for higher level players to start their reasoning (see discussion in Crawford, 2008). Furthermore, our specification is directly related to the  $L0$ -specification in the standard beauty contest. In particular, when  $r = 1$  and signals are public, our game is reduced to a beauty contest game and the two  $L0$  definitions coincide. Given that  $L0$  is non-strategic its behavior should not change as one signal becomes private or as we vary  $r$ . An alternative way to model  $L0$ , which is related to truthful  $L0$  in Crawford and Iriberri (2007), is to assume that the  $L0$  type ignores all strategic aspects of the game (guessing other players' actions) and focuses solely on the nonstrategic aspect of the game (guessing the state). In our setting these two approaches yield the same prediction for higher types' behavior.<sup>2</sup>

According to the standard level- $k$  model, an  $L1$  agent expects that other players are  $L0$  players. This means that an  $L1$  player believes that the average action of other players will be equal to *their own* estimated state:  $\bar{a}_{-i} = \mathbb{E}_{-i} [\theta] = \mu x_{-i}$ . In the setting when  $x_i$  is private  $\mathbb{E}_i [x_{-i}] = \mathbb{E}_i [\theta]$  from which it follows that

$$\mathbb{E}_i [\bar{a}_{-i}] = \mathbb{E}_i [\mu x_{-i}] = \mu (\mu x_i) = \mu^2 x_i.$$

Therefore, an  $L1$  player in the setting with private signals will play

$$a_{L1} = (1 - r) \mu x_i + r\lambda \mu^2 x_i.$$

We use induction to derive the action choice of a level- $n$  agent. Let  $a_{Ln}$  denote the action taken by an  $Ln$  player with private signal  $x_i$ . Then it takes the following linear form  $a_{Ln} = \beta_n x_i$ , where  $\beta_n$  is a coefficient depending on  $r, \lambda$  and  $\mu$ . In particular

$$\beta_0 = \mu \text{ and } \beta_1 = (1 - r) \mu + r\lambda \mu^2. \tag{7}$$

Now consider an  $L(n + 1)$  player with private signal  $x_i$ . Then she expects that other players are  $Ln$  players and

$$\mathbb{E}_i [\bar{a}_{-i}] = \mathbb{E}_i [\beta_n x_{-i}] = \beta_n \mu x_i.$$

Therefore,

$$a_{L_{n+1}} = (1 - r) \mathbb{E}_i [\theta] + r\lambda \mathbb{E}_i [\bar{a}_{-i}] = (1 - r) \mu x_i + r\lambda \mu \beta_n x_i.$$

---

<sup>2</sup>When  $L0$ 's actions are uniformly distributed between the two signals the average  $L0$ 's action will be  $\mathbb{E}_i [\theta] = \frac{1}{2} x_i$ , the exact same number that a (truthful)  $L0$  type would choose.

It follows that

$$\beta_{n+1} = (1 - r)\mu + r\lambda\mu\beta_n,$$

which implies the following difference equation:

$$(\beta_{n+1} - \beta_n) = r\lambda\mu(\beta_n - \beta_{n-1}).$$

Using the initial condition (7), we can solve

$$\beta_n = \frac{(1 - r)\mu}{1 - r\lambda\mu} + \frac{(1 - \lambda\mu)r\mu(r\lambda\mu)^n}{1 - r\lambda\mu}. \quad (8)$$

When signal  $x_i$  is public, by following a similar procedure we can show that an  $Ln$  agent with signal  $(x_i, 0)$  will choose action  $\tilde{\beta}_n x_i$ , where  $\tilde{\beta}_n$  is given by

$$\tilde{\beta}_n = \frac{(1 - r)\mu}{1 - r\lambda} + (1 - \lambda)r\mu \frac{(r\lambda)^n}{1 - r\lambda}. \quad (9)$$

Above we derived level- $k$  predictions under the assumption that subjects are capable of correctly estimating signals received by others. For the setting with private signals we also consider an alternative level- $k$  model where players can not perform appropriate Bayesian updating in estimating  $x_{-i}$ . We call it a naïve level- $k$  model and we assume that subjects are naïve in that they simply think that the other players' private signal is exactly the same as their own. In the experimental part of the paper we will test whether subjects use naïve update or not and the frequency of actions consistent with naïve levels of reasoning. Mathematically, naïve update is equivalent to the case when subjects receive two public signals. Thus, the level- $k$  prediction for naïve update is given by (9). Notice also that in this case if  $\lambda = 1$  then all level- $k$  players will play action  $\mu x_i$  regardless of  $k$ .

It is clear from (8) and (9) that both  $\beta_n$  and  $\tilde{\beta}_n$  are decreasing in  $n$  and converges to our NE predictions given by (5) and (6) as  $n \rightarrow \infty$ . Therefore, we have proved the following result:

**Proposition 1** *All level- $k$  players choose higher actions than the NE prediction.*

When  $\lambda = 1$  the weights put on public and private signals sum up to 1. Therefore, it follows from Proposition 1 that when  $\lambda = 1$  level- $k$  agents will overweight the private signal and underweight the public signal as compared to the theoretical prediction. This has an important implication with regards to the MS model. One of the main results of Morris and Shin (2002) is that the coordination motive forces players to place too much weight on the public signal relative to the weight that would be used by the social planner. As a result, information is not aggregated efficiently and public disclosure of more information could be detrimental to the social welfare. However, Proposition 1 shows that the detrimental effect of public disclosure may be less than predicted by theory if agents are not fully rational. Specifically, level- $k$  players, whether naïve or not, put a higher weight on the private signal — and consequently a lower weight on the public signal — than NE predicts.



## 4 Experimental Design

The design of all treatments in our study is based on the modified MS framework as described above. In this section we explain our experimental implementation of the MS framework as well as similarities and differences across treatments.

### 4.1 Payoff Function and Signals

In all treatments the payoff function of subject  $i$  is given by

$$u_i(a_i, a_{-i}) = 2000 - (1 - r)(a_i - \theta)^2 - r(a_i - \lambda \bar{a}_{-i})^2, \quad (10)$$

where  $a_i$  is the action of subject  $i$ ,  $\theta$  is the true state of the world,  $\bar{a}_{-i}$  is the average of all other subjects' actions,  $\lambda \in [0, 1]$  is the weight put on  $\bar{a}_{-i}$ , and  $r \in [0, 1]$  is the relative importance of matching the weighted average of other investors' actions. Note that negative values of  $u(a_i, a_{-i})$  are possible and so it was publicly announced to participants that negative payoffs would count as 0. Otherwise, subjects may incur a large loss in a single period of the experiment that would be impossible to recover even if subjects receive the maximum of 2000 each period afterwards.<sup>3</sup>

To ensure participants' understanding of the payoff structure we took advantage of the fact that each term had a very simple and intuitive interpretation. We started by verbally explaining that there are three factors that will determine the payoff: mismatching the underlying state, mismatching  $\lambda \bar{a}_{-i}$ , and their relative importance  $r$ . After this was understood, we presented the actual mathematical form, explained the meaning of each term, and went through several numerical examples. Finally, during the actual experiment at the end of each period the second and third terms in (10) were calculated and displayed together with  $a_i$ ,  $\theta$ , and  $\lambda \bar{a}_{-i}$ .

The information available to subject  $i$  was given by two signals:  $y$  and  $x_i$ . The signals as well as state  $\theta$  were generated prior to the experiment according to the following procedure. For each round  $t$ , state  $\theta$  is generated randomly according to a uniform distribution on  $[400, 700]$ . Given  $\theta$ , the signals are independently drawn from a normal distribution  $N(\theta, 3600)$ . Signal  $y$  is public and is the same for all subjects. Signal  $x_i$  can be public or private. In treatments when it is private different subjects in a group observe different signals. When it is public all subjects observe the same signal. Signals and the state were generated in such a way so that each period *all* groups of subjects received the same signals and the underlying state was the same. If, say, members of group 1 received private signals 105, 72, 41 and 36 then in all other groups there would be a member who received signal 105, a (different) member with signal 72 and so on.

After the state and signals were generated, we normalized them by subtracting  $y$  from each of them so the triple  $(\theta, x_i, y)$  becomes  $(\theta - y, x_i - y, 0)$  and the normalized signal  $y$ , therefore, is always 0. Both normalized signals are then displayed on the computer screens and the payoffs are calculated using the normalized state value,  $\theta - y$ . Note that normalized  $x$ -signals and the

---

<sup>3</sup>This can potentially affect the equilibrium prediction since when the maximum of (10) is negative the agent would be indifferent between all actions. However, one can show that this happens only when the two signals are very far apart. In our experiment this happened at approximately 0.1% of all observations.

normalized state could be negative. While the main reason for using the normalization is theoretical and was explained in Section 3 there are also additional benefits. First, it simplifies the environment as it is easier to make a decision with signals 0 and 43 than with signals 529 and 572. Second, this guarantees that subjects know that  $y$  was indeed a public signal. Third, it makes our setting similar to the standard beauty contest setting.

To keep matters simple subjects were not informed about the distributions used for state and signal generation. Subjects were told, however, that the best guess for the state is the average of two signals (see instructions in the Appendix for the exact wording). As one can see from Section 3.2, derivations of levels of reasoning and the equilibrium action do not require knowledge of the distribution as long as one knows how to estimate the state given the two signals.

## 4.2 Treatment and Session Description

There are four aspects in which the MS model differs from the classical beauty contest. First, in the MS model there is private information because agents receive private signals. Second, the goal is divided between guessing  $\lambda$ -average and guessing fundamentals. Third, the action domain is unrestricted. Finally, in the standard MS model  $\lambda = 1$  and in the classical beauty contest  $\lambda < 1$ . The treatments designed for this paper will reflect these differences.

It will be convenient to classify each treatment based on the information structure and value of  $\lambda$ . In total, there are three groups of treatments. In the first group signal  $x_i$  is private and  $\lambda = 1$ . We label this group **Pr-A** as the non-zero signal was priate and the participants must match the average action of other investors. This environment is directly related to the MS model, especially when the domain is unrestricted. In the second group we set  $\lambda = 1/2$  so that subjects need to match  $\theta$  and *one-half* of the average action of their opponents. The latter consideration makes the game related to the  $p$ -beauty contest with  $p = 1/2$ , however, the information is private and not public. We label the group **Pr-H** where the H represents that individuals must now match *one-half* of the average action. Our third group  $\lambda = 1/2$  as in Pr-H but both signals are public. As such only two signals are drawn every period, and it is common knowledge that both signals are public. We label this treatment **Pu-H** as the non-zero signal is now a public signal and subjects need to match *one-half* of the average action. Pu-H is directly related to the beauty contest especially when the domain is restricted. For a fixed information structure and value of  $\lambda$  we will vary values of  $r$  from 0.15 to 0.95 with higher  $r$  corresponding to a higher coordination motive. Finally, for every information structure and  $(r, \lambda)$  pair there is a treatment where the strategy choice is bounded by the two signals, like in BC, and one where it is unrestricted, like in MS.

Given the goal of this paper it is instructive to be more precise regarding the relationship between Pu-H and the beauty contest. First, similar to the beauty contest Pu-H is the game with perfect information. Second, when the choices are restricted to  $[0, x_t]$  it makes Pu-H dominance solvable.<sup>4</sup> Finally, as  $r$  gets closer to 1, the state,  $\theta$ , becomes irrelevant with the only remaining

---

<sup>4</sup>To see this, recall that the best response is given by  $a_i = (1 - r)x_i + r\lambda\bar{a}_{-i}$ . Without loss of generality we can assume  $x_i > 0$ . Because subjects are restricted to choose actions between  $[0, x_i]$ , we can first eliminate actions outside

goal being to match  $\frac{1}{2}\bar{a}_{-i}$ . One notable difference from the BC model is that here *all* subjects, not just the player who is closest to one-half of the average, are paid. However, the tournament aspect of the BC is still retained in that subjects with actions closer to  $\lambda\bar{a}_{-i}$  in Pu-H receive higher payoffs than those farther away.

To sum up there is a group of treatments that is close to the MS model, Pr-A; a group of treatments that is close to the beauty contest, Pu-H; and a group of treatments that is in between the two, Pr-H. Table 1 summarizes the information about the treatments, their mnemonic names, and the number of subjects in each treatment.

	$y$	$x_i$	$\lambda$	Unrestricted Domain	Restricted Domain
Pr-A	0	private	1	19	8
Pr-H	0	private	$\frac{1}{2}$	13	8
Pu-H	0	public	$\frac{1}{2}$	17	25

Table 1: Description of experimental sessions and the number of subjects.

Sessions are based on one of the three designs described. Within each session the information structure, value of  $\lambda$  and restrictions on the domain remained the same with the only variation being due to changes in  $r$ . Each each consists of 6 phases with 10 rounds in each phase, for a total of 60 rounds.<sup>5</sup> Within each phase the value of  $r$  is fixed but  $r$  is different across phases. We use six values of  $r$ : 0.15, 0.3, 0.5, 0.65, 0.8 and 0.95. For each session we use the following order of  $r$  across phases: 0.15, 0.5, 0.8, 0.95, 0.3, and finally 0.65. Thus, in the first phase (first 10 rounds) subjects make decisions with  $r = 0.15$ , while in the second phase (rounds 11-20) subjects make decisions with  $r = 0.5$  and so on. Note that we start with a low value of  $r$ , gradually increase  $r$  until phase four, decrease  $r$  between the fourth and fifth phases, and then increase it again. The choice of a non-monotone sequence of  $r$ 's can help us separate the effect of  $r$  from the effect of learning. For example, if subjects' behavior is similar in phases with  $r = 0.15$  (the first 10 rounds) and  $r = 0.3$  (the fifth ten rounds) then it suggests that this behavior is caused by low  $r$  and not by lack of subject's experience with the environment.

Overall, our design enables us to vary the standard beauty contest setting in the following two directions. First, by changing  $r$  we vary the strength of the coordination motive. This is interesting because games in which the importance of coordination varies can capture a wide range of economic applications such as monetary policy (Morris and Shin, 2002), asset pricing (Allen, Morris and Shin, 2003, Bacchetta and Wincoop, 2005), venture capital (Angeletos, Lorenzoni and Pavan, 2007) and political campaigns (Dewan and Myatt, 2007, 2008). While levels of reasoning are well-defined for any value of  $r$ , one would expect that subjects will focus less on the actions of others as the coordination motive becomes weaker. If this is correct it would suggest that in games

---

of the interval  $[(1-r)x_i/2, (1-r)x_i/2 + r\lambda x_i]$ . Once we do that, we can further eliminate actions outside of the interval  $[(1+r\lambda)(1-r)x_i/2, (1+r\lambda)(1-r)x_i/2 + r^2\lambda^2 x_i]$  and so on. By repeating this procedure we will get a sequence of intervals with length  $r^k\lambda^k x_i$ , and this sequence will shrink to a point, which is NE.

<sup>5</sup>Due to time constraints the Pr-H treatment with restricted domain had only five phases. The last phase (the one with  $r = 0.65$ ) was not conducted.

where coordination is less important or its effect is less obvious subjects will be less likely to follow level- $k$  reasoning.

Second, we introduce private information into the game by making the second signal  $x_i$  private. Private information is prevalent in many economic applications and therefore it is important to understand how well level- $k$  models can explain the data in settings with private information. Indeed, level- $k$  reasonings have been applied to classical settings with private information, such as the winner’s curse in common value auctions and overbidding in private value auctions (see Crawford and Iriberri, 2007a). However, the comparison of level- $k$  model performance between the complete and private information settings, both in absolute and relative terms, has not been studied yet.

### 4.3 Procedures

Sessions were conducted at UNC Charlotte between 2008 and 2010. Subjects were typically undergraduate students, primarily recruited from the business school but not exclusively. Subjects were seated at visually isolated carrels and were forbidden to communicate with other subjects throughout the duration of the experiment. Instructions were read aloud to subjects, and a few minutes were spent discussing how different values of  $r$  could impact the subjects’ loss from mismatching the state  $\theta$  (i.e. the term  $-(1-r)(a_i - \theta)^2$ ) and the loss from mismatching the decisions of other investors (i.e. the term  $-r(a_i - \lambda \bar{a}_{-i})^2$ ). To reinforce this distinction in the actual experiment after each round a payoff screen displayed the loss from mismatching each of these two terms as well as the total payoff.

All subjects were divided into four-person groups which were re-assigned in the beginning of each period. In some sessions we had a number of subjects that was not divisible by 4. In those instances we used the following procedure. First, the computer would form as many groups as possible. The remaining subjects would form an incomplete group that was completed by the decisions of a subject(s) from fully completed groups. When relevant the subject(s) chosen from the fully formed group was the one who observed the private signal different from those observed by members of incomplete group. For instance, if the private signals in a fully completed Pr-H were 105, 72, 41, and 36, and the private signals of an incomplete group were 105, 72, and 41, then a decision from a subject who saw a private signal of 36 would be used to complete the incomplete group. Even though the decision of this randomly chosen subject is used for two groups, that subject will only receive the payoff based on the outcome within her fully formed group.

At the beginning of each round, subjects were shown signals and were asked to submit a decision for  $a_i$ . Depending on the treatment, subjects were informed that either both signals were public signals or one was a public signal and the other was a private signal only observable to that specific subject. When all decisions were submitted,  $\bar{a}_{-i}$  and profit were calculated for each agent. At the end of each round subjects were shown a screen containing their own action choice,  $a_i$ , the true state,  $\theta$ , the average opponent action,  $\bar{a}_{-i}$  and their payoff for the current round.

Subjects’ cash payment is determined as follows. At the end of the experiment one of the six

phases is randomly chosen. A subject’s total payoff during the chosen phase is calculated and converted it into USD by multiplying it by .001. Thus, if a subject earned 10500 during the chosen phase it will become 10.50\$. This is in addition to the 5\$ show-up fee that all subjects received. The average payment to subjects, including the show-up fee, was 15\$ for a 75-90 minute session.

## 5 Results

In this section we analyze subjects’ behavior and study how well it matches NE and level- $k$  ( $Lk$ ) predictions. Given that NE and  $Lk$  actions are linear combinations of a random non-zero signal  $x$  and zero signal  $y$ , they will vary each period even when the treatment and the value of  $r$ , that is the session phase, are fixed. To make results comparable across periods and treatments we normalize the non-zero signal to be 100 and adjust subjects’ actions as well as NE and  $Lk$  predictions accordingly. For example, given action  $a$  and non-zero signal  $x$ , the normalized action is  $a_n = 100 \cdot a/x$  so that action  $a = x/2$  is normalized to 50 and action  $a = x$  is normalized to 100. The interpretation of normalized values is that they represent the percentage weight a particular action or a prediction puts on a non-zero signal.

As mentioned in Section 4, two treatments for each set of parameter values and each information structure were conducted. In one the action domain is restricted to be between the two signals a subject receives and in the other the action domain is unrestricted. Differences in behavior between those two treatments were minimal, and to keep the paper focused we present our findings using the pooled data from the restricted and unrestricted treatments.<sup>6</sup>

### 5.1 Comparing Subjects’ Behavior with NE and Level-k Reasoning

First, we compare subjects’ behavior with NE predictions. For each treatment and each  $r$  we calculate the average normalized action  $a_n$  and plot it on Figure 1 together with the normalized NE prediction. As we see from Figure 1, in all three treatments subjects’ actions are higher than NE predicts. In other words, subjects tend to overweight the non-zero signal which is private in Pr-A and Pr-H and public in Pu-H.<sup>7</sup> Recall that, as we established earlier, overweighting the non-zero signal is consistent with level- $k$  reasoning. Using non-parametric signed rank test we have that most of the time the difference between observed behavior and NE is significant. More precisely, it is significant in all three cases for high values of  $r$ , that is when  $r = 0.8$  and  $r = 0.95$ . In Pr-A the difference is significant for every  $r \neq 0.15$ , in Pu-H for every  $r \neq 0.65$ . As for Pr-H the difference is insignificant for  $r = 0.15$ ,  $r = 0.3$  and  $r = 0.65$ .

Figure 1 shows the average of normalized actions for each treatment. In Table 2 we present the average absolute deviation of actions from NE (in normalized units). As we see deviations are quite substantial in all treatments and especially in treatments with private signals. Notably, the observed behavior was closest to NE when  $r = 0.65$  regardless of the value of  $\lambda$  and the information

<sup>6</sup>Separate results for restricted and unrestricted treatments are available from the authors upon request.

<sup>7</sup>In a setting similar to our Pr-A treatment Cornand and Heinemann (2009) also observed the overweighting of private signals by subjects.

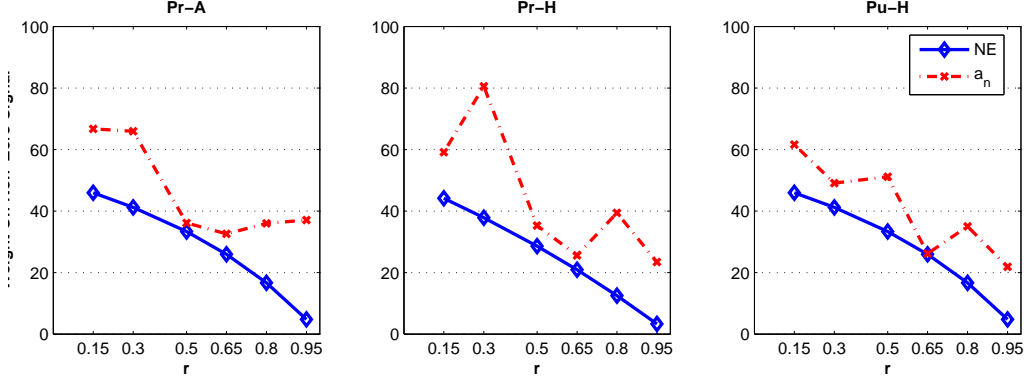


Figure 1: Subjects’ behavior and NE in all treatments. On the  $y$ -axis is the average weight that subjects put on the non-zero signal which is public in  $Pu - H$  and private otherwise. Solid line is NE; and dash-dotted line,  $a_n$ , is the average over normalized actions.

structure. The most likely reason is that  $r = 0.65$  case was the last phase in each session and subjects’ learning could bring them closer to NE. Level- $k$  reasoning is usually thought of as the framework that describes people’s behavior in the beginning of experiments. Therefore, a better performance of NE in the final stage of experimental sessions is not surprising.

r	0.15	0.30	0.50	0.65	0.80	0.95
<b>Pr-A</b>	67.20	53.86	36.88	30.08	46.23	44.76
<b>Pr-H</b>	58.57	86.97	37.14	23.28	44.22	36.93
<b>Pu-H</b>	30.63	16.46	22.43	10.71	18.91	17.59

Table 2: Average absolute deviation of observed behavior from NE across different treatments and phases. The deviation is calculated based on normalized data with non-zero signal normalized to 100. Higher  $r$  means stronger coordination motive.

**Result 1:** *Subjects tend to put a higher weight on the non-zero signal than NE predicts. This is consistent with level- $k$  behavior. Overall, NE performs the best in the last phase of the study with  $r = 0.65$ .*

So far we established that aggregated subjects’ behavior is consistent with level- $k$  predictions in that subjects overweight the non-zero signal. Now we study whether this finding is the result of aggregation and whether it holds on the individual level as well.

Figure 2 visualizes subjects’ behavior and the relationship between observed behavior and level- $k$  predictions. From each of the three groups of treatments we picked three subjects and plotted their choices and levels of reasoning in the phase with  $r = 0.95$ . The choices are denoted by crosses. The solid red line corresponds to  $L1$  and dash-dotted magenta line corresponds to  $L2$ . Dashed black line in treatments with public signals correspond to  $L3$  and dashed green line in treatments with private signals correspond to  $L1naïve$ .

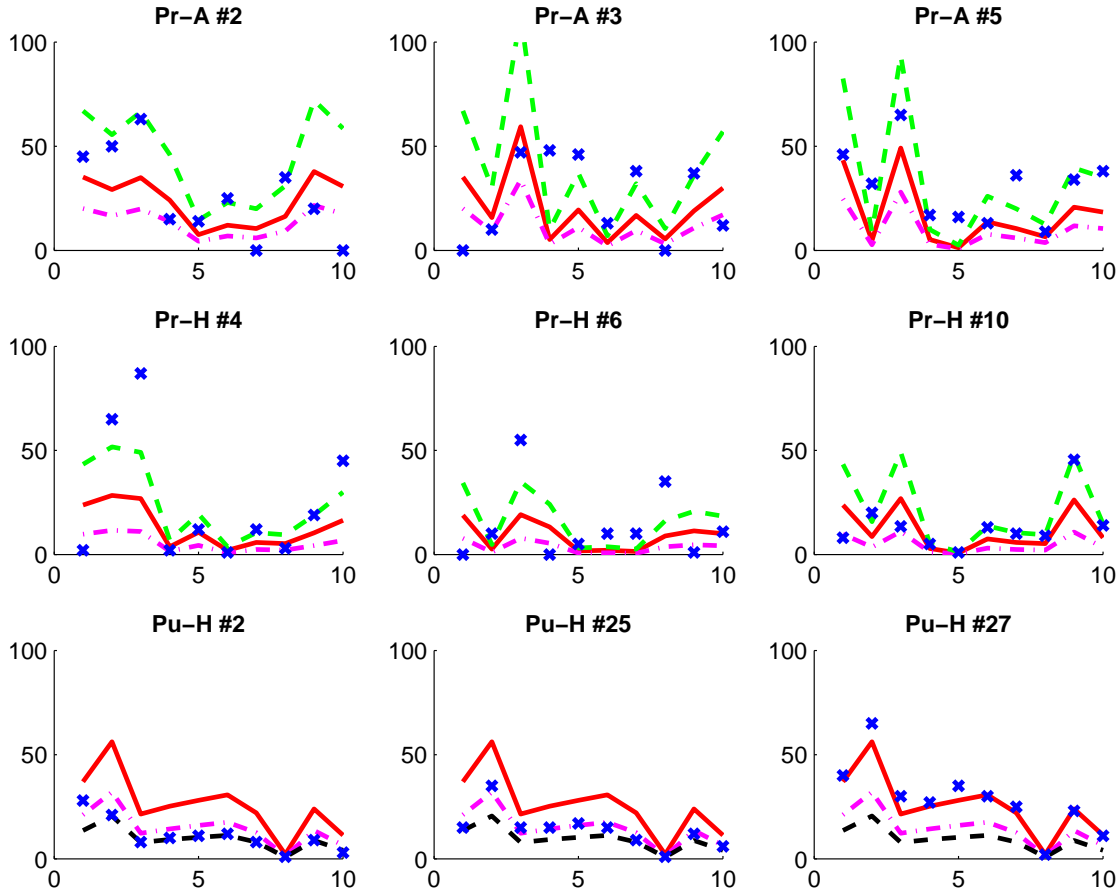


Figure 2: Individual behavior of subjects in phase with  $r = 0.95$ . Crosses correspond to subjects' actions; solid line is  $L1$  and dash-dotted line is  $L2$ . For treatments with public signals dotted line is  $L3$ , for treatments with private signals dotted line is  $L1naïve$ . To increase the scale of images we use absolute values so that all levels and actions are positive.

From Figure 2 we see that some subjects were following a particular level of reasoning in a fairly consistent fashion whereas some subjects were not. In particular, following level- $k$  reasoning was considerably more common in treatments with public signals. Subject 2, for example, was following  $L3$  very consistently. Subject 25 was using  $L2$  and subject 27 seemed to converge to  $L1$  by the second half of the phase. In treatments with private signals, on the other hand, such a pattern was less common. Only subject 10 in  $Pr - H$  seemed to pick choices consistent with  $L1naïve$ ; as for the remaining subjects, their behavior is harder to assign to a particular level of reasoning whether naïve or not.

To quantitatively measure whether actions of a particular subject are consistent with level- $k$  reasoning we use the following criterion. Recall that each experimental session consisted of six 10-period phases with each phase corresponding to a particular value of  $r$ . We divided each phase into two halves: the first five periods and second five periods. Within each group of five periods for each subject we calculated average absolute deviations of subjects' normalized actions from normalized levels of reasoning ( $L1$ ,  $L2$ , and  $L3$ ). We say that a subject's behavior *closely followed*

one of the levels of reasoning during a given half of the phase if two conditions hold. First, the average deviation of subjects' choices from this level was the smallest as compared to other levels. Second, the average deviation from that particular level was less than 10 normalized units.

The reasons why we picked this criterion are as follows. Level- $k$  behavior is often considered as a way to describe subjects' initial responses in which case it should be more pronounced in the beginning of the phase. Furthermore, subjects' behavior can change as the experiment proceeds; in particular, they can switch from one level to another, presumably a more sophisticated one. Separately studying subjects' behavior in the beginning and in the end of a given phase enables us to detect level- $k$  behavior if it is used only in the beginning or if a subject switches levels within a phase. We allow for the error of up to 10 normalized units in order to take into account that even if a particular subject uses level- $k$  reasoning his choices are likely to be biased towards integers, particularly those ending with 0 and 5. Thus it is unlikely that subjects' actions will precisely match a given level of reasoning.<sup>8</sup>

r	0.15	0.30	0.50	0.65	0.80	0.95
<b>Pr-A</b>	12.96	7.41	18.52	24.07	14.81	16.67
<b>Pr-H</b>	7.14	0.00	11.90	34.62	28.57	28.57
<b>Pu-H</b>	11.90	38.10	29.76	60.71	64.29	85.71

Table 3: The frequency with which the first three levels of reasoning were closely followed. We say that a subject's behavior closely followed one of the levels of reasoning during a given half of the phase if two conditions hold. First, the average deviation of subjects' choices from this level was the smallest when compared to other levels. Second, the average deviation was less than 10 (in normalized units).

Table 3 shows the frequency with which the first three levels of reasoning were closely followed in a given treatment. It is calculated as follows. We label decisions made by a particular subject during a particular half of a phase as whether following some  $Lk$  or not. For private treatments we also include naïve levels of reasoning. Then for given  $r$  and information structure we divide the number of half-phases where levels of reasoning were used over the total number of half-phases. Several things can be noticed from Table 3. First, in treatments with public signals the success rate of level- $k$  models tends to be higher. Second, the highest success rate occurs in treatments with two public signals and  $r = 0.95$  where more than 80% of subjects followed some level of reasoning. This suggests that level- $k$  models perform best when there is no private information and the coordination motive is the strongest. The third result, while in some sense being a corollary

<sup>8</sup>As a robustness check we also examined several alternative criteria. We performed the calculations using only the first five periods of each phase (as in Crawford and Iriberri, 2007a) as well as pooling the data from all ten periods of each phase. In addition to having the threshold of 10 normalized units we considered thresholds of 5 and 15 normalized units. We also used actual values instead of normalized ones. In all these cases, the qualitative picture does not change. Level- $k$  models perform better in treatments with public signals and in phases with high  $r$ . Quantitatively, numbers change as compared to Table 3 depending on whether the criterion is more or less favorable to level- $k$  reasoning. If it is more favorable, say because of a higher threshold, then all frequencies are higher. If it is less favorable, say, because of a lower threshold or because we consider all 10 periods of the phase instead of two five-period intervals, then all frequencies are lower.



of the previous two, is worth mentioning separately. In treatments with private signals even when  $r = 0.95$  the success rate of level- $k$  models is relatively low and, in particular, is much lower than in treatments with public signals.

**Result 2:** *In treatments with private signals and in phases with low  $r$  only a few subjects followed levels of reasoning. In treatments with public signals and high values of  $r$  level- $k$  models did the best with the majority of subjects following some level of reasoning.*

The next question is how subjects' behavior evolves over time. First, we study how the predictive power of the level- $k$  model changes within the phase. Given that level- $k$  reasoning is usually viewed as the way to describe subjects' initial responses we measure the performance of level- $k$  models separately in the beginning and in the end of each phase. In Table 4 we calculate percentages of level- $k$  subjects in the first five periods of each phase as well as percentages of level- $k$  subjects in the last five periods of each phase. The criterion for attributing subjects' behavior to a particular level is the same as before. That is the average absolute deviation from this level prediction should be the smallest compared to other levels and should not be more than 10 normalized units.

<b>r</b>	<b>First 5 rounds</b>						<b>Second 5 rounds</b>					
	<b>0.15</b>	<b>0.30</b>	<b>0.50</b>	<b>0.65</b>	<b>0.80</b>	<b>0.95</b>	<b>0.15</b>	<b>0.30</b>	<b>0.50</b>	<b>0.65</b>	<b>0.80</b>	<b>0.95</b>
<b>Pr-A</b>	14.81	11.11	18.52	33.33	11.11	22.22	11.11	3.70	18.52	14.81	18.52	11.11
<b>Pr-H</b>	4.76	0.00	14.29	46.15	23.81	19.05	9.52	0.00	9.52	23.08	33.33	38.10
<b>Pu-H</b>	14.29	40.48	21.43	64.29	50.00	83.33	9.52	35.71	38.10	57.14	78.57	88.10

Table 4: Percentage of subjects who followed closely one of the levels of reasoning in the beginning of the phase (the left table) and in the end of the phase (the right table). Frequencies are computed similarly to Table 3.

Comparing the left and right parts of Table 4 we find no evidence that level- $k$  reasoning was more prevalent in the beginning of the phase. A simple comparison shows that in 9 half-phases out of 18 level- $k$  behavior was more frequent in the beginning, in 2 half-phases frequencies were equal, and in the remaining 7 half-phases it was more frequent in the end. However, in the case of public signals and  $r \geq 0.8$ , which is when level- $k$  reasoning was most common, we see it was used more frequently in the second half. Furthermore, we do not see the evidence for level- $k$  reasoning in the first phases of the experiments which were the phases with  $r = 0.15$  and  $r = 0.5$ . We conclude that our results are inconsistent with the conjecture that level- $k$  reasoning is more likely to be observed in the beginning of the experiment.

Next, we turn our attention to the  $r = 0.8$  and  $r = 0.95$  phases of the treatments with public signals which is where level- $k$  behavior was the most prominent. From Table 4 we see that in these phases level- $k$  reasoning was more pronounced during the second half rather than during the initial five periods. Looking at the evolution of subjects' choices over time we have the following results. Among those subjects who followed some level of reasoning in *both* halves of a phase there were 22 cases (out of 34) when subjects stayed with the same level in both halves, 11 cases when subjects switched to a higher level, and 1 case when a subject switched to lower level (subject 7, phase  $r = 0.95$ , treatment Pu-H). This is consistent with Nagel (1995) who also found that subjects

Subject #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
<b>r=0.8</b>																	
1st half						L1		L2					L1	L1	L3	L1	L1
2nd half	L2	L2		L1	NE	L1	L1	L2	L1	L2		L3	L1	L1	NE	L1	L3
<b>r=0.95</b>																	
1st half	L2	L3	L2	L2		L2	L2	L2	L2	L2	L1	L3	L1	L1	L3	L1	L3
2nd half	NE	L3		L2	L3	L2	L1	L2	L2		L1	NE	L2	L1	L3	L1	L3

Table 5: Subjects' behavior in Pu-H treatment with unrestricted domain in phases with  $r = 0.8$  and  $r = 0.95$ .

tend to adhere to the same level of reasoning throughout the entire study.<sup>9</sup> For those subjects who followed some level in one half of a phase, only 3 subjects did so in the first half while 15 subjects did so in the second half. In other words, it was more likely for subjects to switch to level- $k$  behavior than to abandon it. Similarly, looking at subjects' behavior between phases we see that those subjects who followed some level in the  $r = 0.8$  phase would typically continue doing so, though perhaps using a different level, in the next ( $r = 0.95$ ) phase. Table 5 summarizes the behavior of subjects in the Pu-H treatment with unrestricted domain. For brevity we omit the similar table for 25 subjects from the Pu-H treatment with restricted domain.

Subject #	2	6	8	12	14	17	18	84	85	86	90
<b>r=0.8</b>											
1st half								Ln		Ln	L1
2nd half	NE	NE			Ln				L1	Ln	
<b>r=0.95</b>											
1st half		L2	L1		L1	Ln	Ln	Ln			
2nd half				Ln	Ln		Ln				

Table 6: Subjects' behavior in treatment Pr-A. Only those subjects whose behavior could be attributed to any level are shown. Subject's IDs below 19 are for Pr-A with unrestricted domain and those above 80 for Pr-A with restricted domain.

In treatments with private signals, the picture is considerably less clear. Table 6 shows all participants in the Pr-A treatments whose actions could be attributed to a particular level in at least one half-phase. The difference between Tables 5 and 6 is immediate. Out of 27 subjects in Pr-A treatments there were only 11 who closely followed some level of reasoning during at least one half of at least one phase. That includes two instances when subjects followed NE. Furthermore,

<sup>9</sup>Nagel (1995) finds that subjects' choices decreased over time which one may interpret as the evidence that subjects learn to play with higher levels of reasoning. Nagel argues that this interpretation is false. The declining pattern of choices is not because subjects learned to play with higher levels of reasoning but because subjects adjusted downwards their beliefs about the average action given the outcome of the previous play. Our experiment is different in that every period a new signal – or the support of the beauty contest game – is randomly drawn. Therefore, subjects cannot directly use the information of the average action of the previous play to guess the average action of the current play. This makes learning much slower and more difficult in our setting. As a result we can observe many subjects staying with the same level of reasoning without adjusting subjects' beliefs as in Nagel (1995).

levels of reasoning were not used in a particularly consistent manner. There are no subjects who would follow some level of reasoning throughout both phases, and only three subjects followed some level throughout the entire phase.

**Result 3:** *We do not observe that subjects are more likely to follow level- $k$  reasoning in the beginning of the phase as compared to the end, or vice versa. However, in phases where level- $k$  behavior was most prominent, level- $k$  reasoning was more common in the end.*

**Result 4:** *In treatments with public signals and high  $r$  subjects were more likely to switch to  $Lk$ -behavior and within  $Lk$  behavior they were more likely to switch to higher levels. In treatments with private signals such a consistent pattern is either not observed or is considerably weaker.*

The last question we want to address in this section is how frequently different levels of reasoning were followed by subjects. This information is given in Table 7 where we count how many times subjects followed a particular level of reasoning during a half of the corresponding phase.

	Pr-A				Pr-H				Pu-H			
	L1	L2	Ln	NE	L1(n)	L2(n)	L3n	NE	L1	L2	L3	NE
0.15	1	-	3	3	0(2)	-	-	1	7	-	-	3
0.30	-	1	3	-	-	-	-	-	21	2	-	9
0.50	2	-	7	1	0(2)	-	1	2	14	5	-	6
0.65	3	-	6	4	5(0)	-	-	4	11	10	2	31
0.80	2	-	4	2	2(1)	3(5)	1	-	27	18	5	4
0.95	2	1	6	-	2(2)	1(2)	4	1	29	31	10	2
<b>Total:</b>	10	2	29	10	9(7)	4(7)	6	8	109	66	17	55
<b>Total %:</b>	3.1%	0.6%	9.0%	3.1%	3.8%(3.0%)	1.7%(3.0%)	2.5%	3.4%	21.6%	13.1%	3.4%	10.9%

Table 7: Each entry shows how many times subjects followed a given level within one half of a corresponding phase. In Pr-H treatments columns  $L1(n)$  and  $L2(n)$  show both sophisticated and naïve levels. Numbers for naïve levels are in parenthesis. Numbers in the last row are calculated as a percentage of *all* phase-halves played by *all* subjects within a given treatment.

Several things can be noticed upon inspecting Table 7. First, in treatments with private signals naïve levels were used more often than the “*sophisticated*” ones: 29 versus 12 in Pr-A and 20 versus 13 in Pr-H. Second, in the treatments with public signals as well as in the Pr-H treatment,  $L1$ , regardless of naïvety, was followed most frequently. In Pr-A we cannot separate  $L1n$  from higher naïve levels but we still observe that  $L1$  was followed more often than  $L2$ . Third, in the  $r = 0.8$  and  $r = 0.95$  phases of Pu-H treatments higher levels of reasoning were used more often than in phases with lower  $r$ . In particular, in the  $r = 0.95$  phase level  $L2$  was followed more frequently than  $L1$ . Finally, we see the effect of learning. In the last phase of each treatment, the one with  $r = 0.65$ , NE was followed most often.

**Result 5:** *In all four treatments subjects were more likely to follow the first level of reasoning independent of naïvety. One notable exception is the  $r = 0.95$  phase in treatments with public signals where  $L2$  was most common. In treatments with private signals naïve levels of reasoning were considerably more common.*

Given the popularity of naïve reasoning we conclude this section by studying whether such popularity is due to naïve updating or not. We added a belief-elicitation stage for treatments Pr-A and Pr-H with restricted domain. To minimize the interference between subjects’ responses to belief-elicitation and subjects’ action, the elicitation was conducted at the end of the round, i.e. after the decision was made, and only in the last three rounds of each phase. In other words among 10 rounds for each given value of  $r$ , we added belief elicitation to the end of rounds 8, 9, 10.

The belief elicitation consisted of a single question: “*What was the AVERAGE of private signals observed by other members of your group?*”. While answering the question subjects could see their own private signal. To provide incentives subjects were paid for being close to the correct answer according to the formula  $\max\{100 - (AvgSignal_{-i} - Guess_i)^2, 0\}$ . Denote subject  $i$ ’s guess as  $g_i$  and define  $\alpha_i$  so that  $g_i = \alpha_i \cdot x_i$ . If a subject with signals 0 and  $x_i$  can perform the update correctly then  $\alpha_i = 1/2$ . For naïve updaters  $\alpha_i = 1$ . The average value of  $\alpha$  for all subjects was 0.45 in Pr-A and 0.55 in Pr-H. A  $t$ -test could not reject the hypothesis of correct update with  $p$ -values equal to 0.15 for Pr-A and 0.66 for Pr-H. On the other hand the hypothesis of naïve update was rejected with  $p$ -values of 0.000 in both treatments.

To investigate the issue further we assumed that subjects best responded to their beliefs regarding  $\bar{a}_{-i}$ , which then could be calculated from the chosen actions. Looking separately at subjects who followed a naïve level of reasoning (“naïve” subjects) and at those who did not, we observed the following. First, we did not find evidence of naïve update among “naïve ” subjects. Typically “naïve ” subjects expected the average signal observed by opponents to be around  $x_i/2$ . In fact, “naïve” subjects were more consistent in submitting  $x_i/2$  than the others. Second, for low values of  $r$  (up until 0.5) “naïve” subjects’ median expectation of  $\lambda\bar{a}_{-i}/2$  was  $x_i/2$  both in Pr-A and Pr-H. Thus it seems that following a naïve level of reasoning was not due to naïve update regarding the signals observed by others. Instead, naïve subjects were naïve regarding how the others would act on their information. For example, in Pr-A they would anticipate that their opponents would take actions equal to their signal. They would correctly estimate the opponents’ signal to be centered around  $x_i/2$  and would expect that average opponents’ action to be equal to  $x_i/2$  as well.

## 5.2 Maximum Likelihood Estimation of Level- $k$ and CH Models

Our initial analysis of the data performed in the previous section indicated that the predictive power of level- $k$  reasoning is weaker as we introduce asymmetric information or reduce the weight of the coordination component. In this section we turn to formal statistical analysis of the data to provide further evidence for this finding. We calculate the ML estimations of the following two models: the standard level- $k$  model where  $Lk$  type plays best response to the population that consists entirely of  $Lk - 1$  type, and the cognitive hierarchy model (CH) where  $Lk$  plays the best response to the population that consists of the mixture of lower types  $L0, \dots, Lk - 1$ .

Results of ML estimation of the standard level- $k$  model are given in Table 8. For the estimation we assume that  $L0$  players choose their actions using the uniform distribution and  $Lk$  players play the best response to the  $Lk - 1$  strategy plus an error that is uniformly distributed around the best

<b>Pr-A</b> (1620 obs.)	<b>0.15</b>	<b>0.3</b>	<b>0.5</b>	<b>0.65</b>	<b>0.8</b>	<b>0.95</b>
L1	0.015	0.011	0.052	<b>0.119</b>	<b>0.073</b>	0.018
L2	0.000	0.000	0.036	0.019	0.000	0.003
Ln	0.133	0.105	<b>0.169</b>	<b>0.147</b>	<b>0.113</b>	<b>0.066</b>
NE	0.000	0.000	0.000	0.000	0.013	0.018
LL	<i>-1224.105</i>	<i>-1233.795</i>	<i>-1221.027</i>	<i>-1207.605</i>	<i>-1205.330</i>	<i>-1238.693</i>
<b>Pr-H</b> (1180 obs.)						
L1	0.004	0.001	0.007	0.043	<b>0.133</b>	0.088
L2	0.000	0.012	0.007	0.070	0.000	0.011
L1n	0.067	0.018	0.039	0.000	0.002	<b>0.094</b>
L2n	0.059	0.000	0.000	0.035	0.019	0.000
NE	0.000	0.012	0.014	0.081	0.030	<b>0.097</b>
LL	<i>-951.667</i>	<i>-967.123</i>	<i>-966.540</i>	<i>-581.163</i>	<i>-925.130</i>	<i>-944.501</i>
<b>Pu-H</b> (2520 obs.)						
L1	0.000	<b>0.082</b>	<b>0.076</b>	<b>0.101</b>	<b>0.213</b>	<b>0.291</b>
L2	0.035	0.000	0.018	0.000	<b>0.221</b>	<b>0.353</b>
NE	0.035	0.082	0.049	<b>0.279</b>	<b>0.037</b>	<b>0.101</b>
LL	<i>-1925.36</i>	<i>-1894.41</i>	<i>-1909.39</i>	<i>-1789.14</i>	<i>-1788.41</i>	<i>-1659.33</i>

Table 8: Maximum likelihood estimation of percentage of population that followed  $L1$ ,  $L2$  and  $NE$  in different treatments. Estimates are in bold when they are significantly different from 0 at the 5% level.

response. The estimation is performed using normalized values and the support of the error is set equal to 10 normalized units, i. e. to 10% of the distance between two signals. In the literature the error is usually modeled as having a logistic distribution. We opted for the uniform distribution because subjects are often biased towards “nice” numbers such as integer numbers or those ending with 0 and 5 — something that we also observe in our data. The uniform distribution seems to be a more natural way to capture the error generated by such a bias. In all treatments we directly estimate the shares of  $L1$ ,  $L2$  and  $NE$  types.<sup>10</sup> In addition, in treatments with private signals we estimate shares of naïve types. The share of type  $L0$  is then calculated as one minus the sum of the shares of other types.

The results of ML estimation are consistent with our earlier findings. First, larger values of  $r$  generate larger and, most importantly, significantly positive estimates of shares of different types. Furthermore, in treatments with public signals statistically significant level- $k$  reasoning appear for lower values of  $r$  and the share of population using  $L1$  and  $L2$  are considerably higher than in treatments with private signals. Looking at  $r = 0.95$  and ignoring statistical significance and the  $NE$ -type, we see that in Pr-A only 8.7% used some level of reasoning; in Pr-H it was only 19.3%. For Pu-H this number is 64.4% which is more than 3 times higher than Pr-H and more than seven times higher than in Pr-A. The most prevalent level of reasoning used by participants was Level-1, whether naïve or not. The most notable exception is the Pu-H case with  $r = 0.95$  where  $L1$ ’s share was 29.1% and  $L2$ ’s share was 35.3%.

<sup>10</sup>Adding type  $L3$  does not significantly change the results. It changes estimates slightly for the phase with  $r = 0.95$  only.

Next, we estimate the cognitive hierarchy model (CH) that was introduced in Camerer et al. (2004). The idea behind the CH model is that higher types believe that opponents' population is a mixture of lower types. For example, type  $L2$  believes that some opponents are  $L1$  and others are  $L0$  and best responds accordingly. Camerer et al. assume that there is a correct distribution of different types given by the Poisson distribution with parameter  $\tau$  so that  $\Pr(Lk) = f(k) = \exp(-\tau)\tau^k/k!$ . Each type does not realize that there are players of the same or higher types but it correctly estimates relative proportions of lower types. For example, type  $L2$  will believe that the share of  $L0$  is  $f(0)/(f(0) + f(1))$  and the share of  $L1$  is  $f(1)/(f(0) + f(1))$ .

Pr-A (1620 obs.)	<b>0.15</b>	<b>0.3</b>	<b>0.5</b>	<b>0.65</b>	<b>0.8</b>	<b>0.95</b>
$\tau$	0.093	0.104	0.141	0.378	0.208	0.128
Share of Naïve	0.203	0.206	0.312	0.268	0.225	0.174
LL	-1091.876	-1104.757	-1036.004	-967.282	-1053.433	-1121.754
Pr-H (1180 obs.)						
$\tau$	0.270	0.150	0.187	0.296	0.381	0.489
Share of Naïve	0.953	0.538	0.540	0.105	0.250	0.499
LL	-849.916	-920.093	-904.859	-524.774	-830.011	-799.971
Pu-H (2520 obs.)						
$\tau$	0.169	0.270	0.239	0.658	0.974	2.957
LL	-1800.505	-1698.000	-1741.534	-1693.434	-1410.552	-996.066

Table 9: The ML estimation of the CH model. Higher  $\tau$  implies more evidence of level- $k$  behavior.

The estimation of treatments with public signals is straightforward. As before we assume that each type plays the best response plus the error that is uniformly distributed around the best response. The support of the error is 10 normalized units. In estimation we assumed that the highest type in the population is  $L3$ . For treatments with private signals, we adjust the CH estimation to account for naïve types. First, we assume that both naïve and sophisticated types are unaware of higher types. Second, naïve types are unaware of the sophisticated types. For sophisticated types we considered two alternatives: one when sophisticated types are unaware of naïve types and another one when they are aware. We report only the former and the results for the latter were fairly similar. Finally, we assume that the entire population is divided into two groups: naïve players and sophisticated players. The size of each group is an estimation parameter. Within each group types are distributed according to Poisson distribution with parameter  $\tau$ .

Table 9 shows the results of estimation. A higher  $\tau$  implies that there are larger fractions of subjects who can do higher levels of reasoning. In Pr-A the highest  $\tau$  is actually at  $r = 0.65$  after which it declines. This is consistent with what was seen in Table 8 where the treatment with  $r = 0.65$  leads to the highest coefficients. In Pr-H and Pu-H the estimates of  $\tau$  increase as  $r$  increases especially when  $r = 0.95$ . Except for the  $r = 0.15$  case the  $\tau$  values in Pu-H are considerably higher than in treatments with private signals.

One way to compare the performance of the two level- $k$  models both across the treatments and with each other is to look at the estimated share of  $L0$ . In our estimation procedure any action that was not consistent with a particular level of reasoning was assigned to level 0. Thus higher share

of  $L_0$  implies more observations that cannot be explained by a model. The results are presented in Table 10. In addition to the already well-established fact that level- $k$  models perform the best in public treatments with high  $r$ , we see that the performance of the CH model is considerably better. This is despite the fact that the CH model has fewer parameters than the standard level- $k$  model.<sup>11</sup> For example, in Pu-H with  $r = 0.95$  the CH model leaves unexplained only 6% versus 26% unexplained by the standard level- $k$  model.

Level- $k$	0.15	0.3	0.5	0.65	0.8	0.95
Pr-A	0.85	0.88	0.74	0.72	0.80	0.90
Pr-H	0.87	0.96	0.93	0.77	0.82	0.71
Pu-H	0.93	0.84	0.86	0.62	0.53	0.26
CH	0.15	0.3	0.5	0.65	0.8	0.95
Pr-A	0.91	0.90	0.86	0.65	0.80	0.87
Pr-H	0.74	0.85	0.82	0.72	0.65	0.57
Pu-H	0.83	0.74	0.77	0.47	0.33	0.06

Table 10: Estimates of  $L_0$  by the standard level- $k$  and the CH models.

**Result 6:** *ML estimates confirm that level- $k$  models perform the best when the coordination motive is the strongest and information is public. In the remaining cases level- $k$  models perform poorly and put most of the weight on the non-strategic  $L_0$  type.*

**Result 7:** *In all treatments the CH model fits subjects' behavior better than the standard level- $k$  model, despite the fact that the CH model has fewer parameters than the standard level- $k$  model.*

We conclude the section by comparing the effectiveness of estimated level- $k$  models in predicting the aggregated subject behavior relative to the NE prediction. In Table 11 we calculate the absolute difference (in normalized units) between observed and predicted behavior by NE, level- $k$  model and the CH model. We call these differences prediction errors. The comparison is done at the aggregated level; that is for each treatment we compare the average observed behavior with the average predicted behavior. Differently from NE, both the standard level- $k$  and the CH model assume heterogeneous subjects' behavior which is why we compare the three models only on the aggregated level.

The panel with NE errors shows a picture which is very similar to Figure 1. On the aggregated level NE performs quite well when  $r = 0.65$  and when  $r = 0.5$  and signals are private. In fact,  $r = 0.65$  phase is the only phase where NE consistently outperforms the other two models. As we mentioned above the most likely reason is that  $r = 0.65$  was the last phase and subjects' behavior could converge to NE simply due to learning. For high values of  $r$  the NE error is around 20 normalized units or more which corresponds to 20% of the distance between signals. Both level- $k$  models outperform NE for high  $r$ . This is despite the fact that on the individual level in treatments with private signals and high  $r$  we did not find strong evidence of level- $k$  behavior. A particularly

<sup>11</sup>The CH model has one parameter in treatments with public signals and two in treatments with private signals. The standard level- $k$  model has three parameters in Pu-H, 5 in Pr-H and 4 in Pr-A.

	NE Error					
	0.15	0.30	0.50	0.65	0.80	0.95
Pr-A	20.77	24.77	2.81	6.64	19.30	32.29
Pr-H	14.97	42.72	6.68	4.69	26.89	20.16
Pu-H	15.62	7.91	17.80	0.24	18.31	17.09
	Level- $k$ Error					
	0.15	0.30	0.50	0.65	0.80	0.95
Pr-A	16.77	16.03	12.64	15.09	12.14	11.60
Pr-H	9.64	31.43	13.29	16.58	4.56	15.86
Pu-H	11.86	0.43	3.18	15.49	3.33	4.31
	CH Error					
	0.15	0.30	0.50	0.65	0.80	0.95
Pr-A	16.99	16.56	12.60	13.33	10.95	10.51
Pr-H	10.03	31.85	12.13	17.57	1.80	14.81
Pu-H	12.16	0.88	3.81	15.73	1.67	0.98

Table 11: Average deviation (in normalized units) of the observed behavior from the predicted behavior in NE and level- $k$  models.

striking result is that the error is as little as 1.8 in the  $r = 0.8$  phase of Pr-H; however, it appears to be a result of aggregation. Finally, compared with level- $k$  predictions the NE errors for  $r = 0.95$  case are much higher in treatments with private signals.

**Result 8:** *With the exception of  $r = 0.65$  phase, both level- $k$  models predict the aggregated behavior better than NE. Their advantage is particularly pronounced in treatments with strong coordination motives.*

## 6 Concluding Remarks

The goal of this paper is to determine the setting in which level- $k$  thinking most appropriately describes subjects' behavior. To do that we generalize the classical beauty contest setting by using a modified Morris and Shin (2002) framework that allows us to introduce private information and vary the strength of the coordination motive. Having the experimental design based on the MS model generates an environment that is more complex than the one typically used in the level- $k$  literature. Despite this complexity we confirm the finding in the existing beauty contest literature that level- $k$  models are indeed successful in predicting subject behavior when the game setting is close to the classical beauty contest, that is, when information is symmetric and the coordination motive is strong. Moreover, most subjects choose their levels of reasoning consistently in the sense that they either adhere to one particular level or switch to higher levels. However, as we move away from the classical setting, subjects are less likely to follow level- $k$  reasoning. In particular, only a handful of subjects play according to level- $k$  reasoning and those who do tend to use it in a rather inconsistent manner.

We conjecture that the reason for these results is as follows. When the coordination motive weakens, the behavior of other players becomes less important; as such, subjects are less likely



to try to predict it. This is true regardless of whether information is symmetric or not. The introduction of private information into the model weakens level- $k$  behavior even further since the task of predicting the beliefs and actions of opponents becomes considerably more complex. For example, in the  $p$ -beauty contest with  $p = 1/2$ ,  $L1$  logic can be summarized in the following simple phrase: people will just pick actions randomly between 0 and 100 so the average action will be 50 and so I should play 25. In contrast, in the setting with private information the same  $L1$  logic becomes more complicated since subjects do not know the range from which others are choosing and have to estimate it. Given the increased complexity of level- $k$  reasoning, participants may rely on a different rule of thumb in settings with private information. For example, as our analysis shows subjects can take naïve approach regarding how other agents use their information. The identification of the exact rule of thumb subjects used in the experiment is an important research question, and we leave it for future research.

## 7 Appendix. Instructions for Treatment Pr-H

### Welcome to a decision-making study!

#### Introduction

Thank you for participating in today's study in economic decision-making. These instructions describe the procedures of the study, so please read them carefully. If you have any questions while reading these instructions or at any time during the study, please raise your hand. *At this time I ask that you refrain from talking to any of the other participants.*

#### General Description

This study consists of 60 rounds, time permitting. In each round all participants (including you) have the role of investors. All participants are divided into groups with 4 investors in each group. The division is random and will be re-done in the beginning of each round. You and the 3 other investors in your group can invest some amount of experimental currency in a particular project. Your task is to decide how much you would like to invest into this project. Returns on your investment will be determined by the amount that you invest ( $a_{you}$ ) and by the following two factors:

- the project's quality  $q$ ;
- one-half of the average investments made by others:  $\frac{1}{2} \cdot a_{average} = \frac{1}{2} \cdot \frac{a_1 + a_2 + a_3}{3}$ ;

**Example:** Assume that the other three investors in your group invested 150, 200 and 250. The average amount invested by the others is  $a_{average} = 200$ . One-half of the average then is  $\frac{1}{2} \cdot 200 = 100$ .

At the time when you make decisions you will **NOT** know either of these two factors. You will not know one half of the average amount invested by others,  $\frac{1}{2} \cdot a_{average}$ , because other participants are making their decisions at the same time as you. You will not know  $q$  because you must make your investment decision before  $q$  is revealed. Therefore, you will need to decide how much to invest based on the information that will be made available to you.

#### Information. Signals.

In the beginning of each round you and all other investors in your group will receive two signals that will provide you with information about the project's quality. Both signals are randomly drawn given the project's quality  $q$ . Because signals are randomly drawn it is impossible to precisely predict  $q$  given the signals. However, they will give you an idea of a range where  $q$  might be. The Table below shows to you how signals should be interpreted.

First, to make calculations easier for you one signal is always set equal to 0. Second, given the two signals that you will see the best guess of  $q$  will be simply the average of the two signals. Because of the randomness it is unlikely that  $q$  will ever be precisely equal to the average of the two signals. The last two columns in the table give you an idea of how precise your guess is. You see that in two cases out of three, i.e. with probability  $2/3$ , the quality,  $q$ , will be at most 40 away from the average and with probability 95% the quality will be at most 80 away from the average.

Signal 1	Signal 2	The best guess of $q$	With prob. $2/3$ $q$ will be in	With prob. 95% $q$ will be in
0	s	$(0+s)/2$	$(0+s)/2 \pm 40$	$(0 + s)/2 \pm 80$

**Example 1:** Assume that you received two signals 0 and 100. Then the best guess of the project quality would be  $(0 + 100)/2 = 50$ . With probability of  $2/3$  you can conclude that the project quality will be between  $10(= 50 - 40)$  and  $90(= 50 + 40)$  and with probability 95% the project quality will be between -30 and 130. In the remaining 5% of the cases the quality will be outside of the  $[-30, 130]$  interval.

### Guessing one-half of the average

In the previous section we explained how to guess  $q$  given the information that you will receive (the two signals). However, your profit will also depend on how well you can guess one-half of the average amount invested by other investors in your group. The decisions of other investors are decisions made by humans and therefore there is no precise theory that will tell you where one-half of the average will be.

Therefore, your best option would be to try to predict how much the other investors are going to invest given their information. Here is what you know and what you don't know about the information available to other investors in your group:

- They receive two signals, just like you do;
- You know the first signal that everyone receives. It is 0. All investors in your group will have 0 as the first signal.
- You do NOT the second signal that they receive. The second signal is a private signal. It means that you cannot see private signals received by other investors. It also means that they cannot see the private signal that you receive.
- You DO know that private signals of other investors are generated in the same way as your private signal. Most importantly that they are also centered around the project's quality  $q$ .

Use your knowledge about the information that other investors have to predict how much they will invest. Based on that you can form your guess of one-half of the average investment.

## Your Profit and Cash Payments

Your profit will be calculated as follows. In the beginning of each round you will be given 2000 experimental points. From this amount we will deduct points when your action does not match the project's quality. We will also deduct points when your action does not match *one-half* of the average investments made by others. Your final profit will be calculated by the following formula:

$$Payoff = 2000 - (1 - r)(a_{you} - q)^2 - r \left( a_{you} - \frac{1}{2}a_{average} \right)^2.$$

The first term says that your investment will bring you at most 2000. The second term determines your loss from mismatching the project's quality  $q$ . The third term determines your loss from mismatching *one-half* of the average investments made by others.

It is possible that the project quality and one-half of the average investment will be two different numbers. In this case parameter  $r$  measures the **relative importance** of matching the investments of others versus matching the quality. A lower  $r$  means matching the quality is more important. Relative importance will be changed every 10 rounds.

The following two examples are used to illustrate how  $r$  impacts your payoff. While you will submit decisions for these two examples they are for illustrative purposes and will not impact your payment.

**Example:** Let  $r = 0.15$  so that it is *more important to match the quality*. Let quality,  $q$ , be 10, and  $a_{average}$  be 120. At your computer terminal, please submit an action of 30 now. If your action,  $a_{you}$ , is 30 then your loss from mismatching the quality is  $(1 - 0.15) \cdot (30 - 10)^2 = 340$ . Your loss from mismatching one-half of the average investments is  $0.15 \cdot (30 - 60)^2 = 135$ . You see that your mismatch of the average investment is larger than the mismatch of quality, but your losses from mismatching the quality are higher. Your total profit is  $2000 - 340 - 135 = 1525$ .

**Example:** Now assume that  $r = 0.8$  so that it is *more important to match the investments of others*. As before assume that  $q = 10$  and  $a_{average} = 120$ . Thus everything is the same as in the example above except for  $r$ . Again, please submit an action of 30 now. Your loss from mismatching the quality is  $(1 - 0.8) \cdot (30 - 10)^2 = 80$  and your loss from mismatching the average investment is much higher and is equal to  $0.8 \cdot (30 - 60)^2 = 720$ . Your total profit is  $2000 - 80 - 720 = 1200$ .

The profit that you made in each round will be converted into cash by the following procedure. The study lasts for 60 rounds. In the end of the study we will openly and randomly choose a sequence of 10 rounds: either from round 1 to round 10, or from round 11 to round 20 and so on. Your cash earnings will be equal to the total profit that you earned during these 10 rounds times 0.001. This is in addition to the \$5 that you receive as a show-up fee. For example, if round 21 to 30 is chosen and you earned 10000 during these rounds your cash payoff will be:  $10000 \cdot 0.001 + 5 = \$15$ . If in a particular round you make a negative profit it will count as 0.

## Summary

The study consists of 60 rounds, time permitting. In the beginning of each round, the computer will generate the project quality  $q$  and randomly determine 3 other investors who will be in your group. Computer will also generate two signals for each participant. The first signal — zero — will be the same among all participants. The second signal will be private. It means that you cannot see the signals received by other investors, and they cannot see the second signal received by you.

Your task is to submit an amount that you would like to invest. After you and all other members of your group enter their decisions, the computer will calculate and display your profit in that particular round. Your profit will be determined based on how well you guessed the project's quality and how well you guessed one-half of the average investment made by others. In the end of the study we will take the profit you made in a randomly chosen sequence of 10 rounds and will convert it into cash payment.

## References

- [1] Allen, Franklin, Stephen Morris and Hyun Song Shin (2006). “Beauty Contests and Iterated Expectations”, *Review of Financial Studies*, 19, 161-177.
- [2] Angeletos, George-Marios, Guido Lorenzoni and Alessandro Pavan (2007). “Wall Street and Silicon Valley: A Delicate Interaction.” Working paper, MIT.
- [3] Angeletos, George-Marios, and Alessandro Pavan (2007). “Efficient Use of Information and Social Value of Information.” *Econometrica*, 75(4), 1103–42.
- [4] Bacchetta, Philippe, and Eric van Wincoop (2005). “A Theory of the Currency Denomination of International Trade”, *Journal of International Economics* 67, 295-319.
- [5] Bosch-Domenech, Antoni, Jose Garcia Montalvao, Rosemarie Nagel, and Albert Satorra (2002). “One, Two, (Three), Infinity, . . . : Newspaper and Lab Beauty-Contest Experiments.” *American Economic Review*, 92, 1687–1701.
- [6] Camerer, Colin, Teck-Hua Ho, and Juin-Kuan Chong (2004). “A Cognitive Hierarchy Model of Games.” *Quarterly Journal of Economics*, 119, 861–898.
- [7] Cornand, Camille, and Frank Heinemann (2009). “Measuring agents’ Overreaction to Public Information in Games with Strategic Complementarities.” Work in progress, Department of Economics and Management, Technical University Berlin.
- [8] Costa-Gomes Miguel, Vincent Crawford and Bruno Broseta (2001) “Cognition and behavior in normal-form games : An experimental study”, *Econometrica* vol. 69, no 5, pp. 1193-1235.
- [9] Costa-Gomes, Miguel, and Vincent Crawford, (2006). “Cognition and Behavior in Two-Person Guessing Games: An Experimental Study”, *American Economic Review*, vol. 96, pp. 1737-1768.

- [10] Costa-Gomes, Miguel, Vincent Crawford, and Nagore Iriberry (2009), “Comparing Models of Strategic Thinking in Van Huyck, Battalio, and Beil’s Coordination Games,” *Journal of the European Economic Association*, forthcoming.
- [11] Crawford, Vincent, and Nagore Iriberry (2007a). “Level-k Auctions: Can a Non-Equilibrium Model of Strategic Thinking Explain the Winner’s Curse and Overbidding in Private-Value Auctions?” *Econometrica*, 75, 1721–1770.
- [12] Crawford, Vincent, and Nagore Iriberry (2007b). “Fatal Attraction: Saliency, Naïvete, and Sophistication in Experimental Hide-and-Seek Games.” *American Economic Review*, 97, 1731-1750.
- [13] Crawford, Vincent (2008) “Modeling Behavior in Novel Strategic Situations via Level- $k$  Thinking”, Third World Congress of the Game Theory Society, 14 July 2008.
- [14] Dewan, Torun, and David Myatt (2007). “Leading the Party: Coordination, Direction and Communication.” *American Political Science Review*, 101(4), 825–843.
- [15] Dewan, Torun, and David Myatt (2008). “The Qualities of Leadership: Communication, Direction and Obfuscation.” *American Political Science Review*, 102(3), 351–368.
- [16] Gneezy, Uri (2005). “Step-Level Reasoning and Bidding in Auctions.” *Management Science*, 51, 1633–1642.
- [17] Ho, Teck-Hua, Colin Camerer, and Keith Weigelt (1998). “Iterated Dominance and Iterated Best Response in Experimental ‘ $p$ -Beauty Contests’.” *American Economic Review*, 88, 947-969.
- [18] Morris, Stephen and Hyun Song Shin (2002). “The Social Value of Public Information.” *American Economic Review* 92 (2002), pp. 1521-1534.
- [19] Nagel Rosemarie, 1995. “Unraveling in Guessing Games: An Experimental Study.” *American Economic Review*, Vol. 85, No. 5, pp. 1313-1326.
- [20] Stahl, Dale and Paul Wilson (1995). “On Players’ Models of Other Players: Theory and Experimental Evidence.” *Games and Economic Behavior*, 10, 218-254.
- [21] Stahl, Dale and Paul Wilson (1994). “Experimental Evidence on Players’ Models of Other Players.” *Journal of Economic Behavior and Organization*, 25, 309–327.