A Bayesian Level-k Model in n-Person Games

Teck-Hua Ho, So-Eun Park, and Xuanming Su *

March 19, 2015

In standard models of iterative thinking, players choose a fixed rule level from a fixed rule hierarchy. Nonequilibrium behavior emerges when players do not perform enough thinking steps. Existing approaches, however, are inherently static. This paper introduces a Bayesian level-k model, in which players perform Bayesian updating of their beliefs on opponents' rule levels and best-respond with different rule levels over time. In this way, players exhibit sophisticated learning. We apply our model to experimental data on p-beauty contest and price matching games. We find that it is crucial to incorporate sophisticated learning to explain dynamic choice behavior.

Over the past few decades, accumulating evidence from laboratory and field experiments has shown that players' actual behavior may significantly deviate from the equilibrium solution in a systematic manner. That is, despite nontrivial financial incentives, subjects frequently do not play the equilibrium solution. For example, the first round offer in sequential bargaining games often neither corresponds to the equilibrium offer nor is accepted immediately (Binmore, Shaked and Sutton, 1985; Ochs and Roth, 1989); centipede games often proceed to intermediate stages rather than end immediately (McKelvey and Palfrey, 1992); and many players do not choose the unique iterative dominance solution in dominance solvable games (Nagel, 1995; Stahl and Wilson,

*Ho: University of California, Berkeley and National University of Singapore, 2220 Piedmont Avenue, Berkeley, CA 94720, hoteck@haas.berkeley.edu. Park: University of British Columbia, 2053 Main Mall, Vancouver, British Columbia V6T1Z2 Canada, soeun.park@sauder.ubc.ca. Su: University of Pennsylvania, 3620 Walnut Street, Philadelphia, PA 19104, xuanming@wharton.upenn.edu. Authors are listed in alphabetical order. Direct correspondence to any of the authors. We thank Colin Camerer, Vince Crawford, David Levine, and seminar participants at National University of Singapore, University of British Columbia and University of California, Berkeley for their helpful comments.

1994, 1995; Ho et al., 1998; Costa-Gomes et al., 2001; Bosch-Doménech et al., 2002; Costa-Gomes and Crawford, 2006; Costa-Gomes and Weizsäcker, 2008).

To explain nonequilibrium behavior in games, cognitive hierarchy (Camerer et al., 2004) and level-k (Nagel, 1995; Stahl and Wilson, 1994, 1995; Costa-Gomes et al., 2001; Costa-Gomes et al., 2009; Costa-Gomes and Crawford, 2006; Crawford, 2003; Crawford and Iriberri, 2007b) models have developed; for a comprehensive review, see Crawford et al., 2013. These models assume that players choose rules from a well-defined discrete rule hierarchy. The rule hierarchy is defined iteratively by assuming that a level-k rule best-responds to a lower level rule (e.g., level-$(k-1)$ where $k \geq 1$). The level-0 rule can be specified a priori either as uniform randomization among all possible strategies or as the most salient action derived from the payoff structure.[1] Let us illustrate how to apply these models using two examples: p-beauty contest and traveler's dilemma games.

- In the p-beauty contest game, n players simultaneously choose numbers ranging from 0 to 100. The winner is the player who chose the number closest to the *target number*, defined as p times the average of all chosen numbers, where $0 < p < 1$. A fixed reward goes to the winner, and it is divided evenly among winners in the case of a tie. This game is dominance-solvable and the unique number that survives the iterative elimination process is 0.[2] To apply

[1]The level-k and cognitive hierarchy models have also been applied to study games with asymmetric information such as zero-sum betting, auctions and strategic information disclosure (e.g., Camerer et al., 2004; Crawford and Iriberri, 2007a.; Brocas et al., 2010; Brown et al., 2010; Östling et al., 2011). Crawford et al. (2009) employ level-k models to determine the optimal design for auctions when bidders are boundedly rational. Goldfarb and Yang (2009) and Goldfarb and Xiao (2011) apply the cognitive hierarchy model to capture heterogeneity in the abilities of firms or managers, and use it to predict a firm's long-term success in oligopolistic markets.

[2]To illustrate, let us assume $p = 0.7$. Strategies between 70 and 100 are weakly dominated by 70 and thus are eliminated at the first step of elimination of dominated strategies. Likewise, strategies between 49 and 70 are eliminated at the second step of elimination as they are weakly dominated by 49. Ultimately, the only strategy that survives the iterative elimination of dominated strategies is zero, which is the unique iterative dominance solution of the game.

the standard level-k model, assume that the level-0 rule randomizes between all possible choices between 0 and 100 and thus chooses 50 on average. Best-responding to the level-0 opponents, the level-1 rule seeks to hit the target number in order to maximize the payoff. Specifically, the level-1 rule chooses $x^* = p \cdot \frac{x^*+(n-1)\cdot 50}{n}$ as a best response to its level-0 opponents, taking into account her own influence on the group mean.[3] Proceeding iteratively, a level-k rule will choose $\left(\frac{p\cdot(n-1)}{n-p}\right)^k \cdot 50$, which converges to 0 as k increases.

- In the traveler's dilemma game, two travelers whose identical antiques were damaged by an airline were promised compensation according to the following scheme. Both travelers were asked to privately declare the value x_i of the antique as any integer dollar between \$2 and \$100. They were then compensated the lower declared value, $\min_i x_i$. In addition, the traveler who declared the lower value is rewarded with \$2 (i.e., he receives $\min_i x_i + 2$), while the other traveler who declared a higher value is penalized with \$2 (i.e., he receives $\min_i x_i - 2$). If both give the same value, they receive neither reward nor penalty (i.e., they both receive $\min_i x_i$). This traveler's dilemma game can also be interpreted as a price matching (i.e., lowest price guarantee) game between two firms selling an identical product. The firm charging the lower price is rewarded with customer satisfaction while the firm charging the higher price is punished with customer dissatisfaction.[4] In this game, the unique iterative dominance solution is 2; in general, it is the lower corner of the strategy space.[5] To apply the level-k model, we as-

[3]Some prior research (e.g., Nagel, 1995) ignores one's own influence on the group average by simply assuming that the level-1 rule chooses $p \cdot 50$ and the level-k rule chooses $p^k \cdot 50$. This approximation is good only when n is large. For example, when $n = 100$, the level-1 rule's exact best response is 34.89, which is close to 35.

[4]Price matching is common in the retail industry. Marketing researchers (Jain and Srivastava, 2000; Chen et al., 2001; Srivastava and Lurie, 2001) suggest that firms which promise to charge the lowest price are perceived negatively if customers find out that the firms' prices are not the lowest.

[5]At the first step of elimination of dominated strategies, 99 and 100 are eliminated since they are dominated by 98. At the second step of elimination, 97 and 98 are eliminated being dominated by 96. Hence, the strategy that ultimately survives the iterative elimination of

sume that the level-0 rule naïvely declares the true value, say 51. Then, the level-1 rule chooses the payoff maximizing strategy, which is $50 = 51 - 1$, receiving a payoff of 52. Proceeding iteratively, the level-k rule chooses $\max\{51 - k, 2\}$, which converges to the equilibrium 2 as k increases.

In both games, nonequilibrium behavior emerges when players do not perform enough thinking steps. Beyond these two examples, level-k models have also been used to explain nonequilibrium behavior in applications including auctions (Crawford and Iriberri, 2007b), matrix games (Costa-Gomes et al., 2001), and signaling games (Brown et al., 2012).[6]

However, standard level-k models cannot account for dynamic trends in observed data in repeated games. In laboratory experiments, subjects' choices are initially far from equilibrium but move closer to it over time (Ho et al., 1998; Capra et al., 1999; Goeree and Holt, 2001). Existing models posit that players' rule levels are fixed, suggesting that if a player is level-k, she will choose rule k in every round. As a result, these models predict a *fixed* distribution of choices as the proportion of players choosing each rule remains unchanged over time. Figures 1 and 2 compare data from the first and last rounds of two laboratory experiments: p-beauty contests games from Ho et al. (1998) and generalized price matching games that we freshly conducted. (We shall discuss this generalized price matching game in detail in Section 2 below.) As in the standard price matching game, the equilibrium in the generalized price matching game is the lower corner of the strategy space, which is 80 in this experiment. As shown, subjects' choices move closer to equilibrium over time.[7]

dominated strategies is 2. In a similar fashion, it can be shown that the unique equilibrium solution of traveler's dilemma with strategy space $[L, U]$ is L.

[6] For a comprehensive review, see Crawford et al. (2013).

[7] In Ho et al. (1998), 2% of the choices were iterative dominance solution plays in the first round and the proportion grew to 13% in the last (10th) round. In the generalized price matching game that we conducted, 10% of the choices were iterative dominance solution plays in the first round and the proportion grew to 86% in the last (10th) round.

For both games, a Kolmogorov-Smirnov test rejects the null hypothesis that the distributions of choices in the first and last rounds are identical ($p < 0.001$). As a consequence, any model seeking to capture this shift in behavior over time needs to be dynamic in nature.

[INSERT FIGURE 1 and 2 HERE.]

One approach to capturing dynamic behavior in games is to incorporate learning. Notably, the experience-weighted attraction (EWA) learning model (Camerer and Ho, 1999; Ho et al., 2007) unifies two streams of adaptive learning models that seemed rather disparate previously: reinforcement learning, where players reinforce only actions chosen by themselves, and belief learning models, where players best-respond based on beliefs formed by past actions chosen by themselves as well as other players. The EWA model thus provides a natural way to incorporate dynamics into existing level-k models. Specifically, consider level-0 players who change their choices over time according to EWA. As long as past actions approach the equilibrium, choices of these adaptive learning level-0 players as well as higher level players will converge to the equilibrium. Thus, by incorporating adaptive learning, level-k models predict convergence towards equilibrium, explaining the main feature of data.

Nevertheless, this approach still assumes that players' rule levels remain fixed over time. The fixed rule levels constrain each rule k's adaptation over time in a specific way. To illustrate, let us return to the p-beauty contest. Suppose the level-0 player chooses 50 in the first round and 25 in the last round. Hence, the level-k player chooses $\left(\frac{p\cdot(n-1)}{n-p}\right)^k \cdot 50$ in the first round and $\left(\frac{p\cdot(n-1)}{n-p}\right)^k \cdot 25$ in the last round. If we scale the level-k player's choice in the last round by $\frac{50}{25}$ (i.e., ratio of the level-0 player's first round choice to her last round choice), we see that the level-k player's first-round and scaled last-round choices become identical. Put differently, upon incorporating adaptive learning, existing level-k models make a sharp prediction: the distributions of first-round and *scaled* last-round choices are identical. We can readily test this sharp prediction using data from p-beauty contest and generalized price matching games. In both games, a Kolmogorov-Smirnov test rejects the null hypothesis that the first-round and scaled last-round choices come from the same distribution ($p < 0.001$). Figures 3 and 4 show the distributions of the first-round, last-round, and scaled last-round choices in p-beauty contest and generalized price matching games, respectively. Contrary to the prediction of existing level-k models, the distribution of scaled last-round choices is to the left of that of first-round choices in both games, suggesting that players choose higher level rules in the last round than in the first round and that adaptive learning models under-predict the speed of convergence towards equilibrium.

[INSERT FIGURE 3 and 4 HERE.]

A sensible way to account for the systematic shift in players' scaled choices is to allow players to become more sophisticated (i.e., change their rule levels) over time. Specifically, in our model, players are assumed to form beliefs about what rules opponents are likely to use, update their beliefs after each round via Bayesian updating based on their observations, and best-respond to their updated beliefs in the subsequent round. Consequently, the distribution of players' rule levels changes over time. To the extent that players move up the rule hierarchy over time, this Bayesian level-k model will be able to explain the leftward shift of the scaled last round choices.

In this paper, we propose a model that generalizes the standard level-k model. This model holds three key features:

1. Our model is general and applicable to any n-person game that has a well-defined rule hierarchy.
2. Our model nests existing nonequilibrium models such as EWA learning,

imitation learning and level-k models, and may serve as a model of equilibration process.

3. Players' rule levels are dynamic in that players perform Bayesian updating of their beliefs about opponents' rule levels and thus may adopt different rules over time.

We apply our proposed model to experimental data from two dominance-solvable games: p-beauty contest and generalized price matching. Estimation results show that our model describes subjects' dynamic behavior better than all its special cases. We find that it is crucial to incorporate sophisticated learning even after other existing nonequilibrium models are introduced.

The rest of the paper is organized as follows. Section 1 formulates the general model and its special cases. Section 2 applies the model to both p-beauty contest and generalized price matching games, and presents empirical results from the best fitting model and its special cases. Section 3 concludes.

1 Bayesian Level-k Model

1.1 Notation

We consider the class of n-player games. Players are indexed by i ($i = 1, 2, \ldots, n$). Player i's strategy space is denoted by S_i, which can be either discrete or continuous. Players play the game repeatedly for a total of T rounds. Player i's chosen action at time t is denoted by $x_i(t) \in S_i$. The vector of all players' actions at time t excluding player i's is denoted by $\mathbf{x}_{-i}(t) = (x_1(t), \ldots, x_{i-1}(t), x_{i+1}(t), \ldots, x_n(t))$. Player i's best response function given her prediction of opponents' actions is denoted by $BR_i(\cdot) : (S_1 \times \cdots \times S_{i-1} \times S_{i+1} \times \cdots \times S_n) \rightarrow S_i$.

1.2 Rule Hierarchy

Players choose rules $k \in \mathbb{Z}_+^0 = \{0, 1, 2, \ldots\}$ from iteratively-defined discrete rule hierarchies, and choose actions corresponding to these rules. Each player has her own rule hierarchy, and each rule hierarchy is iteratively defined such that a level-$(k+1)$ rule is a best response to the opponents' level-k rules (Stahl and Wilson, 1994; Nagel, 1995; Ho et al., 1998; Crawford et al., 2013). In symmetric n-player games, all players' rule hierarchies are identical, but they may not be in general.

Let $k_i(t)$ be player i's chosen rule and $a_i^{k_i(t)}(t) \in S_i$ be the action corresponding to rule $k_i(t)$ under player i's rule hierarchy at time t. Below we discuss how player i chooses $k_i(t)$ and how $a_i^{k_i(t)}(t)$ varies over time as the game evolves.

1.3 Belief Draw and Best Response

Players have beliefs about what rule levels their opponents will choose at time t. Denote player i's belief at the end of time t by $\mathbf{B}_i(t) = (B_i^0(t), B_i^1(t), B_i^2(t), \ldots)$, where $B_i^k(t)$ refers to the probability that player i's opponents will choose rule k in the next time period. Based on $\mathbf{B}_i(t)$, player i conducts a *single* draw $d_i(t+1)$ from $\{0, 1, 2, \ldots\}$ with probability $B_i^{d_i(t+1)}(t)$ at the beginning of time $t+1$.[8] Player i then conjectures that *all* her opponents will choose rule $d_i(t+1)$ at time $t+1$. Put differently, player i predicts that all her opponents j ($\forall j \neq i$) will choose the action $a_j^{d_i(t+1)}(t+1)$ that corresponds to rule $d_i(t+1)$ at time $t+1$.

Player i then best-responds to this prediction by choosing:

[8] In a 2-person game, a single draw assumption is reasonable. In an n-player game, players might conduct multiple draws, perhaps one draw per opponent. If this is indeed the case, each player conducts a separate draw for each opponent and each draw comes from a separate distribution. Prior research suggests that this is unlikely to be true. In fact, opponents' choices are found to be often perfectly correlated (Ho et al. (1998)) and players tend to treat all their opponents as an aggregate player.

$$\begin{aligned} x_i(t+1) &= a_i^{d_i(t+1)+1}(t+1) \\ &= BR_i(a_1^{d_i(t+1)}(t+1), \ldots, a_{i-1}^{d_i(t+1)}(t+1), a_{i+1}^{d_i(t+1)}(t+1), \ldots, a_n^{d_i(t+1)}(t+1)) \end{aligned}$$

Hence, player i's chosen rule is exactly one level higher than her drawn belief of opponents' rule level. Since the lowest rule level players can draw is 0, players always choose rules of level 1 or higher.

1.4 Bayesian Belief Updating

In the Bayesian level-k model, players update their beliefs over time. Specifically, at the end of round t ($t \geq 1$), player i observes her opponents' actions, $\mathbf{x}_{-i}(t)$, and infers what *common* rule level corresponds to these actions in round t, and updates her belief accordingly.

Let player i's initial prior belief about her opponents' rule levels be $\mathbf{B}_i(0) = (B_i^0(0),\ B_i^1(0), B_i^2(0), \ldots)$ before the game begins.[9] We operationalize this initial prior using a set of rule counts $\mathbf{N}_i(0) = (N_i^0(0),\ N_i^1(0), N_i^2(0), \ldots)$ such that $B_i^k(0) = \frac{N_i^k(0)}{\sum_k N_i^k(0)}$. Belief updating is then equivalent to updating these rule accounts based on observed chosen rules. Let player i's rule counts at the end of round t be $\mathbf{N}_i(t) = (N_i^0(t), N_i^1(t), N_i^2(t), \ldots)$ where $N_i^k(t)$ denotes player i's cumulative rule count for rule k. Then, $B_i^k(t) = \frac{N_i^k(t)}{\sum_k N_i^k(t)}$.

After the game play in time t, upon observing all other players' actions $\mathbf{x}_{-i}(t)$, player i infers the probability each rule level could give rise to these observed actions $\mathbf{x}_{-i}(t)$ and updates her rule counts accordingly. Specifically, the probability that $\mathbf{x}_{-i}(t)$ results from a rule k, $P_{-i}^k(t)$, is given by:

$$P_{-i}^k(t) = \frac{B_i^k(t-1) \cdot \prod_{j \neq i} f(x_j(t), a_j^k(t))}{\sum_k B_i^k(t-1) \cdot \prod_{j \neq i} f(x_j(t), a_j^k(t))} \tag{1.1}$$

[9] For common priors, simply assume $\mathbf{B}_i(0) = \mathbf{B}(0),\ \forall i$.

where $f(x, \mu)$ denotes the probability density or mass function of observing the choice x given the predicted choice μ. As a consequence, player i's cumulative rule count for a rule k at the end of time t becomes:

$$N_i^k(t) = \rho \cdot N_i^k(t-1) + (n-1) \cdot P_{-i}^k(t)$$

where ρ is a memory decay factor. These updated rule counts, $\mathbf{N}_i(t)$, yield a posterior belief $B_i^k(t) = \frac{N_i^k(t)}{\sum_k N_i^k(t)}$.

Note that this updating process is consistent with Bayesian updating involving a multinomial distribution with a Dirichlet prior (Fudenberg and Levine, 1998; Camerer and Ho, 1999; Ho and Su, 2013).

Let us emphasize how this memory decay factor (ρ) and the rule counts ($\mathbf{N}_i(t)$) interplay with each other. Let $N_i(0)$ be the sum of player i's rule counts before the game starts. In other words, $N_i(0) = \sum_k N_i^k(0)$ and hence $N_i^k(0) = N_i(0) \cdot B_i^k(0)$. We observe that $N_i(0)$ captures the stickiness of prior belief $\mathbf{B}_i(0)$. That is, $N_i(0) = \infty$ implies that the prior belief is extremely sticky and the belief updating process has no influence on the initial prior at all, i.e., $N_i^k(t) = N_i^k(0) = N_i(0) \cdot B_i^k(0),\ \forall t$. On the contrary, $N_i(0) \approx 0$ implies a weak prior, and belief updating is mainly driven by observed rule levels.

Note that players' chosen rules are not fixed and can vary over time. At each time t, player i's rule level depends on her draw, which is moderated by her belief $\mathbf{B}_i(t-1)$; this belief is then updated to $\mathbf{B}_i(t)$ using the observed choices $\mathbf{x}_{-i}(t)$ after time t. In this regard, players are sophisticated and change their rule levels in response to their opponents' chosen rules in the past.

1.5 Level-0 Rule Dynamics

We incorporate adaptive learning by allowing level-0 players to change their choices over time. Since a rule hierarchy is iteratively defined, choices of all higher level players are affected as well.

We introduce heterogeneity to the types of players. Specifically, we let a player be a level-0, adaptive learning player throughout the game with an α probability; and be a belief-updating, sophisticated player throughout the game with a $1-\alpha$ probability. Put differently, with an α probability, player i adopts rule 0 in every round and thus chooses $a_i^0(t)$ in round t. On the other hand, with a $1-\alpha$ probability, player i undertakes sophisticated learning. That is, she starts the game with an initial prior belief, updates her belief after each round, conducts a single draw to predict which rule others will choose, and best-responds accordingly. Note that by setting $\alpha = 1$, our model reduces to a pure adaptive learning model.

Incorporating adaptive learning into the Bayesian level-k model serves two purposes. First, it unifies the existing wide stream of research on adaptive learning since any adaptive learning model can be integrated as level-0 players' behavior in our model. Second, it helps us to empirically control for the potential existence of adaptive learning and hence tease out the respective roles of adaptive and sophisticated learning in describing players' choice dynamics.

1.6 Model Summary

The Bayesian level-k model ($BL_k(t)$) exhibits the following key features:

1. (Belief Draw and Best Response) At the beginning of round t, player i draws a belief $d_i(t)$ from $\{0, 1, 2, \ldots\}$ with probability $B_i^{d_i(t)}(t-1)$. Player i then conjectures that all her opponents will choose rule $d_i(t)$, i.e., $\hat{\mathbf{x}}_{-i}(t) = (a_1^{d_i(t)}(t), \ldots, a_{i-1}^{d_i(t)}(t), a_{i+1}^{d_i(t)}(t), \ldots, a_n^{d_i(t)}(t))$. Consequently, player i best-responds in round t by choosing $x_i(t) = a_i^{d_i(t)+1}(t) = BR_i(a_1^{d_i(t)}(t), \ldots, a_{i-1}^{d_i(t)}(t), a_{i+1}^{d_i(t)}(t), \ldots, a_n^{d_i(t)}(t))$.
2. (Belief Updating) At the end of round t, player i's belief about her opponents' rule levels may change. Specifically, rule counts are updated af-

ter round t to: $N_i^k(t) = \rho \cdot N_i^k(t-1) + (n-1) \cdot P_{-i}^k(t)$ where $P_{-i}^k(t) = B_i^k(t-1) \cdot \prod_{j \neq i} f(x_j(t), a_j^k(t)) / \left(\sum_k B_i^k(t-1) \cdot \prod_{j \neq i} f(x_j(t), a_j^k(t))\right)$.

3. (Adaptive Level-0 Players) Level-0 players exhibit adaptive learning behavior, i.e., $a_i^0(t)$ changes over time.

$BL_k(t)$ extends existing level-k models in two significant ways. First, it allows players to update their beliefs about opponents' rule levels via Bayesian updating in response to the feedback players receive. Thus, as the game proceeds, a player may choose a rule with different probabilities over time. Specifically, after each round, player i updates her belief that her opponents choose a rule k by inferring the probability that opponents' actions have resulted from that rule. Therefore, belief updating differs across players. From this individually updated belief distribution, each player draws a rule in each period, uses it to guess what rule her opponents are likely to choose, and best-responds by choosing the rule one-level higher.

Second, $BL_k(t)$ incorporates adaptive learning by allowing level-0 players to exhibit adaptive learning behavior. $BL_k(t)$ can accommodate any general specification, including EWA learning and its variants. Since a rule hierarchy is iteratively defined, the behavior of all higher level players (i.e., the rule hierarchy) is influenced by these adaptive learning level-0 players.

1.7 Special Cases

$BL_k(t)$ consists of three broad classes of models as discussed below.

1.7.1 Static Level-k with Adaptive Level-0 Players ($L_k(t)$)

If we suppress belief updating in $BL_k(t)$ by setting $N_i(0) = \infty$, $\forall i$, we obtain $L_k(t)$. $L_k(t)$ not only captures adaptive learning but also allows for the $1 - \alpha$ fraction of higher level players to iteratively best respond to level-0 learning dynamics. However, $L_k(t)$ will predict a different learning path than that of $BL_k(t)$ because the former fails to capture sophisticated learning.

Ho et al. (1998) propose such a model with adaptive level-0 players for p-beauty contests. In their model, rule 0 adapts over time, inducing an adaptive rule hierarchy, whereas higher level thinkers do not adjust their rule levels. In this regard, their model is a special case of $BL_k(t)$ with $N_i(0) = \infty, \forall i$.

By restricting $\alpha = 1$, $L_k(t)$ reduces to a broad class of adaptive learning models such as EWA learning and imitation learning. The EWA learning model (Camerer and Ho, 1999; Ho et al., 2007) unifies two separate classes of adaptive learning models: 1) reinforcement learning and 2) belief learning models (including weighted fictitious play and Cournot best-response dynamics). The weighted fictitious play model has been used to explain learning behavior in various games (Ellison and Fudenberg, 2000; Fudenberg and Kreps, 1993; Fudenberg and Levine, 1993). As a consequence, $BL_k(t)$ should describe learning in these games at least equally well.

$L_k(t)$ can also capture imitation learning models, which borrow ideas from evolutionary dynamics. Specifically, players are assumed to mimic winning strategies from previous rounds. These models have been applied to two- and three-player Cournot quantity games, in which varying information is given to players (Hück, Normann, and Oechssler, 1999; Bosch-Domenech and Vriend, 2003). Imitation learning has been shown to work well in the low information learning environment.

Note that adaptive learning models do not allow for higher level thinkers, who perform iterative best responses to level-0 learning dynamics. As a result, $L_k(t)$ is more general than the class of adaptive learning models.

1.7.2 Bayesian Level-k with Stationary Level-0 Players (BL_k)

By setting $a_i^0(t) = a_i^0(1)$, $\forall i, t$, $BL_k(t)$ reduces to BL_k. In BL_k, players update their beliefs and sophisticatedly learn about opponents' rule levels, while keeping level-0 choices fixed over time. As a result, this class of models only captures sophisticated learning. Ho and Su (2012) adopt this approach and generalize the standard level-k model to allow players to dynamically adjust their rule levels. Their model is shown to explain dynamic behavior in centipede and sequential bargaining games quite well. Nonetheless, their model assumes that the strategy space and rue hierarchy of a player are equivalent. BL_k addresses this drawback by relaxing this equivalence assumption and hence is applicable to any n-person games.

1.7.3 Static Level-k Models (L_k)

If $a_i^0(t) = a_i^0(1)$ and $\mathbf{B}_i(t) = \mathbf{B}_i(0)$, $\forall i, t$, then $BL_k(t)$ reduces to L_k, where both the level-0's choices and higher level thinkers' beliefs are fixed over time. L_k nests the standard level-k model as a special case, which has been used to explain nonequilibrium behavior in applications such as auctions (Crawford and Iriberri, 2007b), matrix games (Costa-Gomes et al., 2001), and signaling games (Brown et al., 2012). As a result, our model will naturally capture the empirical regularities in the above games.

There is an important difference between L_k and the standard level-k model in the existing literature. The standard level-k model prescribes that each player is endowed with a rule at the beginning of the game and the player *always* plays that rule throughout the game. Thus, any choice of a player that does not correspond to her prescribed rule is attributed to errors. On the other hand, our L_k model prescribes that each player is endowed with a static *belief distribution* and she makes an i.i.d draw in each round from this distribution.

This implies that a player may draw different rules over time even though the belief distribution is static. As a consequence, a player's choice dynamics over time are interpreted as heterogeneity in i.i.d draws and hence as rational best responses, instead of errors in her choices. Note that if each player i's initial prior is restricted to a point mass, L_k becomes the standard level-k model.

Note that L_k further reduces to the iterative dominance solution in dominance-solvable games if we restrict $\alpha = 0$, $B_i^{\infty}(0) = 1$ and $B_i^k(0) = 0$, $\forall k < \infty$, $\forall i$ (i.e., all players have a unit mass of belief on rule ∞, which corresponds to the iterative dominance solution). Since $BL_k(t)$ nests the iterative dominance solution as a special case, we can use our model to evaluate the iterative dominance solution empirically. Moreover, as the next proposition shows, $BL_k(t)$ converges to the iterative dominance solution if rule 0 converges to it or if all players are higher level thinkers whose learning dynamics satisfy a set of conditions. (See online Appendix for the proof.)

Proposition 1. *Let x_i^* be player i's unique iterative dominance solution in a dominance-solvable game. Then,* $\lim_{t\to\infty} \mathbb{E}\left[x_i(t)\right] \to x_i^*$, $\forall i$, *if:* (1) $\lim_{t\to\infty} a_i^0(t) \to x_i^*$, $\forall i$, *or* (2) $\lim_{t\to\infty} a_i^0(t) \not\to x_i^*$ $\exists i$ *and the following conditions* (a)-(d) *are met:* (a) $\alpha = 0$; (b) $\forall i$, $\rho < 1$ *and* $N_i(0) < \infty$; (c) $\forall i$, $\mathbf{B}_i(0)$ *follows a uniform distribution*; (d) $\forall i, t$, $f(\cdot, \mu) > 0$, $\inf_{k,k'} \frac{f(a_i^k(t), a_i^k(t))}{f(a_i^k(t), a_i^{k'}(t))} > 1, \forall k, k'$, $\sup_{k,k'} \frac{f(a_i^k(t), a_i^k(t))}{f(a_i^k(t), a_i^{k'}(t))} < \infty, \forall k, k'$, *and* $f(a_i^k(t), a_i^{k_1}(t)) = f(a_i^k(t), a_i^{k_2}(t))$ *for all mutually distinct* k, k_1 *and* k_2.

2 Empirical Results

In this section, we estimate the Bayesian level-k model and its special cases using the two games described above: p-beauty contest and generalized price matching. Below, we describe our data, estimation methods, likelihood function, estimated models and estimation results.

2.1 Data

2.1.1 p-Beauty Contest

We use data from p-beauty contest games collected by Ho et al. (1998). The experiment consisted of a 2×2 factorial design, where $p = 0.7$ or 0.9 and $n = 3$ or 7. Players chose any number from $[0, 100]$. The prize in each round was \$1.5 for groups of size 3 and \$3.5 for groups of size 7, keeping the average prize at \$0.5 per player per round. A total of 277 subjects participated in the experiment. Each subject played the same game for 10 rounds in a fixed matching protocol.[10] For $p = 0.7$, there were 14 groups of size 7 and 14 groups of size 3; for $p = 0.9$, there were 14 groups of size 7 and 13 groups of size 3. Thus, we have a total of 55 groups and 2770 observations.

2.1.2 Generalized Price Matching

Before describing the generalized price matching game, we first provide our rationale for generalizing the standard price matching game. In the standard price matching game, the player choosing the lowest price is rewarded with a strictly positive customer goodwill, and thus the best response is to undercut the lowest-priced opponent by the smallest amount feasible in the strategy space, say $\epsilon = \$0.01$. This strategic interaction is operationalized by traveler's dilemma described in the introduction (with $\epsilon = \$1$). As a consequence, the rule hierarchy is defined in terms of steps of size ϵ and hence is independent of other game parameters such as the size of the reward. Therefore, we develop a generalized price matching game, where a rule hierarchy varies with the game parameters. In this way, we can accommodate changes in the rule hierarchy that may affect choice dynamics.

[10] Reputation building is unlikely because the p-beauty contest game is a constant-sum game where players' interests are strictly opposed.

In the generalized price matching game, n firms (players) engage in price competition in order to attract customers. They simultaneously choose prices in $[L, U]$ for a homogenous good. Firms are indexed by i ($i = 1, 2, \ldots, n$). Firm i's choice of price at time t is denoted by $x_i(t)$. Let the choice vector of all firms excluding firm i be $\mathbf{x}_{-i}(t) = (x_1(t), \ldots, x_{i-1}(t), x_{i+1}(t), \ldots, x_n(t))$ and let $\underline{\mathbf{x}}_{-i}(t) = \min\{\mathbf{x}_{-i}(t)\}$. Firm i's payoff function of the generalized price matching game at time t is then given by:

$$\Pi_i(x_i(t), \mathbf{x}_{-i}(t)) = \begin{cases} \underline{\mathbf{x}}_{-i}(t) - R \cdot s & \text{if } \underline{\mathbf{x}}_{-i}(t) + s < x_i(t) \\ \underline{\mathbf{x}}_{-i}(t) - R \cdot |x_i(t) - \underline{\mathbf{x}}_{-i}(t)| & \text{if } \underline{\mathbf{x}}_{-i}(t) < x_i(t) \leq \underline{\mathbf{x}}_{-i}(t) + s \\ x_i(t) & \text{if } x_i(t) = \underline{\mathbf{x}}_{-i}(t) \\ x_i(t) + R \cdot |x_i(t) - \underline{\mathbf{x}}_{-i}(t)| & \text{if } \underline{\mathbf{x}}_{-i}(t) - s \leq x_i(t) < \underline{\mathbf{x}}_{-i}(t) \\ x_i(t) + R \cdot s & \text{if } x_i(t) < \underline{\mathbf{x}}_{-i}(t) - s \end{cases} \quad (2.1)$$

where $R > 1$ and $s > 0$ are constants.

[INSERT FIGURE 5 HERE.]

Figure 5 illustrates how a player's payoff (the y-axis) depends on the minimum of her opponents' choices (the x-axis). As Eq. (2.1) indicates, there are five possible cases. In all cases, a player's payoff is the lowest price offered by all players (due to price matching) as well as an additional goodwill term (which may be positive or negative depending on the relative position of the player's price). If the player's choice is lower than the minimum of opponents' choices, the player receives positive customer goodwill because she uniquely offers the lowest price. This positive customer goodwill increases linearly at rate R as the player's own choice gets lower. Since $R > 1$, payoff increases as the player's choice decreases because the increase in goodwill more than compensates for the decrease in price. However, once the goodwill reaches its cap of $R \cdot s$, it stops increasing and stays constant. Thus, if the player continues to drop its price, her payoff will start to decrease. On the other hand, if the player's own choice is higher than the minimum of opponents' choices, she receives negative customer goodwill because her price is not the lowest. The magnitude of negative customer goodwill increases linearly in the player's own choice. However, once the goodwill reaches its cap of $-R \cdot s$, its magnitude stops increasing and stays constant. If the player's own choice and the minimum of her opponents' choices are identical, she receives neither positive nor negative customer goodwill and the payoff is simply the lowest price.

There are several implications of the payoff structure of this game. Note that player i's maximum possible payoff is $\min\{\mathbf{x}_{-i}(t)\} + (R-1) \cdot s$, which is attained when $x_i(t) = \min\{\mathbf{x}_{-i}(t)\} - s$, assuming that $x_i(t)$ is above the lowest possible choice L. In other words, player i's best response to her opponents' choices, $\mathbf{x}_{-i}(t)$, is: $x_i^*(t) = \max\{L, \min\{\mathbf{x}_{-i}(t)\} - s\}$. Therefore, we may interpret $s > 0$ as the *step size* by which a best-responding player undercuts her opponents' choices, and $R > 1$ as the *rate of change* of customer goodwill around the lowest price offered by competitors in the market. In addition, when $s \to 0$, $R \to \infty$, and $R \cdot s$ stays constant (e.g., $s = 1/R$ and $R \to \infty$), the game converges to the standard price matching game where the size of goodwill is always fixed regardless of the relative magnitude among choices.

This generalized price matching game is dominance solvable and its unique iterative dominance solution is the lower corner of strategy space, L. Let us illustrate with $[L, U] = [80, 200]$, $R = 1.5$ and $s = 10$. Choices above $U - s = 190$ are dominated by 190 and are eliminated at the first step of iterated elimination of dominated strategies. In the second step, choices above 180 are eliminated. Proceeding in this fashion, only $L = 80$ remains after 12 steps.

Now let us define the rule hierarchy. A generalized price matching game with the strategy space $[L, U]$, $R > 1$ and $s > 0$ has a well-defined rule hierarchy, where a rule k is defined by: $a^k(t) = \max\{L, a^0(t) - s \cdot k\}$.[11]

[11]Players' rule hierarchies are identical in symmetric games. Hence, we suppress the subscript i.

Since the generalized price matching game had not been studied previously, we conducted a new experiment to collect data. The experiment consisted of a 2x2 factorial design with $n = (3, 7)$ and $s = (5, 10)$. We set $R = 1.5$ for all experimental conditions. A total of 157 students from a major university in Southeast Asia participated in the experiment, with 7 groups in the experimental condition of $n = 3$ and $s = 10$, and 8 groups in each of the three remaining experimental conditions. Subjects chose any number from $[80, 200]$. Each subject played the same game for 10 rounds in a random matching protocol.[12] In each round, subjects accumulated points according to the payoff function in Eq. (2.1). After all 10 rounds, these points were converted to dollars at a rate of \$0.01 per point. The equilibrium payoff in all experimental conditions was fixed at \$8. In addition, all participants were paid a \$5 show-up fee.[13] The full set of instructions is given in the online Appendix.

2.2 Estimation Methods

In empirical estimation, we normalized the data as in Ho et al. (1998) so that each round's data has equal influence on the estimation.[14] Since both games are symmetric, we assumed a common prior belief distribution among the players, i.e., $\mathbf{B}(0) = \mathbf{B}_i(0),\ \forall i$ and $N(0) = N_i(0),\ \forall i$.[15]

We assumed that the probability density function $f(x, \mu)$ in Eq. (1.1) follows a normal distribution with mean μ and standard deviation σ, i.e.,

[12] We used a random matching protocol to avoid reputation building since the price matching game has outcomes that are Pareto-superior to the equilibrium outcome. Data shows convergence towards equilibrium, which means that reputation building was indeed thwarted. However, 9.8% of all choices were 200, suggesting that some players still tried reputation building albeit unsuccessfully. We excluded these choices in the estimation.

[13] The prevailing exchange rate at the time of the experiment was US\$1 = \$1.20.

[14] Generally, variance drops over time as convergence takes place. As a result, if data is not normalized, later rounds' data will have more influence on the estimation. Specifically, $x_i(t)$ and $a_i^{k_i(t)}(t)$ were normalized to $(x_i(t) - \mu(t))/s(t)$ and $(a_i^{k_i(t)}(t) - \mu(t))/s(t)$ where $\mu(t)$ and $s(t)$ are the sample mean and sample standard deviation at time t, respectively.

[15] In the belief updating process, as in Camerer and Ho (1999), we restrict $N(0) \leq 1/(1-\rho)$ so that the belief updating obeys the law of diminishing effect. That is, the incremental effect of later observations diminishes as learning equilibrates.

$f(x, \mu) = \mathcal{N}(x, \mu, \sigma)$ where $\mathcal{N}(x, \mu, \sigma)$ denotes the normal density with mean μ and standard deviation σ evaluated at x.[16] The standard deviation σ was empirically estimated. We also estimated a separate standard deviation, σ_0, for the choices of level-0 players, in order to investigate whether these non-sophisticated thinkers' choices are more noisy.[17]

Next, we assumed a common level-0 behavior across players since the game is symmetric, i.e., $a_i^0(t) = a^0(t), \ \forall i$. Further, we assumed that level-0 players are imitation learners who mimic past periods' winning actions that yielded the highest payoff, i.e., $a^0(t)$ follows imitation learning dynamics. Specifically, let $\omega(t)$ be the winning action with the highest payoff in time t that is actually chosen by one or more players, and let $\omega(0)$ be the a priori winning strategy before the game starts. Then, the rule 0 at time t corresponds to a weighted sum of the rule 0's choice at time $t-1$ and the most recent winning action at time $t-1$. Formally, $a^0(t)$ is defined by:

$$\begin{aligned} a^0(1) &= \omega(0) \\ a^0(t) &= \delta \cdot a^0(t-1) + (1-\delta) \cdot \omega(t-1) \quad (t \geq 2) \end{aligned}$$

where δ is the weight given to the rule 0's choice in the previous round. Thus, δ captures the rate at which a player decays past winning actions. Note that when $\delta = 0$, level-0 players completely decay history after each round and simply repeat the most recent winning strategy. Hence, a level-1 thinker who best-responds to this captures Cournot best-response dynamics. On the other hand, when $\delta = 1$, level-0 players never incorporate observed game plays and their choices remain unchanged, i.e., $a^0(t) = a^0(1) = \omega(0), \ \forall t$.[18]

[16] Subjects' actual choices were always discrete; in p-beauty contest games, choices were always integers and in generalized pricing matching games, choices never had more than one decimal point. Therefore, we discretized the strategy space, $f(x, \mu)$ and the likelihood function given below into intervals of 1 for the prior and 0.1 for the latter in our estimation.

[17] Since we used normalized data, $\mathcal{N}(x, \mu, \sigma_0)$ and $\mathcal{N}(x, \mu, \sigma)$ become close to the uniform distribution for σ_0 and σ big enough across both games. Thus, we bounded σ_0 and σ from above by 20 in the estimation. Estimation results were not sensitive to increasing the bound.

[18] We also fit the models with $a^0(t)$ specified as the weighted fictitious play (Fudenberg

We devoted the first round data entirely to calibrating parameter estimates for the a priori winning strategy at time 0, $\omega(0)$, and the prior belief distribution, $\mathbf{B}(0)$.[19] Calibrating the first round data serves two purposes. First, it helps control for the initial heterogeneity in belief draws before any learning takes place. Second, it mitigates under- or over-estimating the speed of learning. For instance, choices too far from the equilibrium may be misattributed to slow adaptive learning by level-0 players, i.e., the model may wrongly predict slow convergence of $a^0(t)$ towards the equilibrium in order to explain those choices. Moreover, in extremely fast converging games, the prior belief distribution may be estimated to be skewed towards higher rule levels in order to explain the high proportion of equilibrium choices overall. This may result in underestimating the speed of belief updating, i.e., overestimating the stickiness parameter $N(0)$ or the memory decay factor ρ in belief updating.

We used the maximum likelihood estimation method with the MATLAB optimization toolbox. To ensure that the maximized likelihood is indeed the global maximum, we utilized multiple starting points in the optimization routine.[20] The 95 % confidence intervals of parameter estimates were empirically constructed using the jackknife method. These confidence intervals are reported in parentheses below parameter estimates in Tables 1 and 2.

---

and Levine, 1998). The weighted fictitious play fits poorly. The results are available upon request from the authors.

[19] Without imposing any parametric assumptions on the belief distribution, we used as many free parameters as the rules that fit in the strategy space. Let this number of rules be K. Note that $K = \infty$ in our p-beauty contest games since rule ∞ corresponds to the equilibrium. Hence, in the estimation, we instead used K large enough ($K = 50$) and the estimation results were not sensitive to having a higher number of rules.

We used the maximum likelihood estimation method. For notational consistency, let us treat the first round data as data at time 0, i.e., $\{x_i(0)\,|\,\forall i\}$ refers to the first round data used for calibration purposes only. Then, the log-likelihood of observing the first round data is given by: $\sum_i \log\left(\sum_{k=0}^{K} B^k(0) \cdot \mathcal{N}(x_i(0), a^k(0), \tilde{\sigma})\right)$ where $a^0(0) = \omega(0)$ and $a^k(0)$ $(k \geq 1)$ is defined iteratively. Here, $\tilde{\sigma}$ is the error term used only for the first round estimation. The estimated $\hat{\omega}(0)$, $\hat{\mathbf{B}}(0)$ and $\hat{\tilde{\sigma}}$ are described in footnotes in Tables 1 and 2.

[20] We randomly selected 20 starting points in the parametric space.

2.3 Likelihood Function

We define likelihood function where each player i's choice probability of choosing $x_i(t)$ at time t is based on her actual interactions with her matched opponents up to time $t-1$.[21] Let $\Theta = (\alpha, \delta, \sigma_0, N(0), \rho, \sigma)$ and let $\overline{\mathbf{x}}(t)$ be the history of actions chosen by *all* players up to time t, i.e., $\overline{\mathbf{x}}(t) = \{x_i(t') \mid \forall t' \leq t \text{ and } \forall i\}$. Further, let $L_i^0(x_i(t) \mid \Theta, \omega(0), \mathbf{B}(0), \overline{\mathbf{x}}(t-1))$ be the probability that a *level-0* player i chooses $x_i(t)$ at time t, and $L_i^h(x_i(t) \mid \Theta, \omega(0), \mathbf{B}(0), \overline{\mathbf{x}}(t-1))$ be the probability that a *higher-level* player i chooses $x_i(t)$ at time t, given Θ, $\omega(0)$, $\mathbf{B}(0)$, and $\overline{\mathbf{x}}(t-1)$. Since each player i's type (i.e., level-0 or higher-level) is fixed throughout the game in our model, the likelihood of observing player i's entire data, $L_i(x_i(1), \ldots, x_i(T) \mid \Theta, \omega(0), \mathbf{B}(0))$, is given by:[22]

$$\begin{aligned}
&L_i(x_i(1), \ldots, x_i(T) \mid \Theta, \omega(0), \mathbf{B}(0)) \\
&= \alpha \cdot \prod_{t=1}^{T} L_i^0(x_i(t) \mid \Theta, \omega(0), \mathbf{B}(0), \overline{\mathbf{x}}(t-1)) + (1-\alpha) \cdot \prod_{t=1}^{T} L_i^h(x_i(t) \mid \Theta, \omega(0), \mathbf{B}(0), \overline{\mathbf{x}}(t-1)) \\
&= \alpha \cdot \prod_{t=1}^{T} \mathcal{N}(x_i(t), a^0(t), \sigma_0 \mid \Theta, \omega(0), \mathbf{B}(0), \overline{\mathbf{x}}(t-1)) \\
&\quad + (1-\alpha) \cdot \prod_{t=1}^{T} \left(\sum_{k=0}^{\infty} B_i^k(t-1 \mid \Theta, \omega(0), \mathbf{B}(0), \overline{\mathbf{x}}(t-1)) \cdot \mathcal{N}(x_i(t), a^{k+1}(t), \sigma \mid \Theta, \omega(0), \mathbf{B}(0), \overline{\mathbf{x}}(t-1)) \right)
\end{aligned}$$

That is, player i's likelihood function is the weighted sum of conditional likelihoods depending on whether she is level-0 or higher-level. The log-likelihood of observing entire data, $LL(\overline{\mathbf{x}}(T) \mid \Theta, \omega(0), \mathbf{B}(0))$, is then given by:

$$LL(\overline{\mathbf{x}}(T) \mid \Theta, \omega(0), \mathbf{B}(0)) \;=\; \sum_{i=1}^{n} \log\left(L_i(x_i(1), \ldots, x_i(T) \mid \Theta, \omega(0), \mathbf{B}(0))\right)$$

[21] This method is called the "closed-loop" approach. Another candidate is the "open-loop" approach, which is to simulate all such interactions in the population. Thus, in the open-loop approach, the choice probability of $x_i(t)$ is identical for all players ($\forall i$), whereas in the closed-loop approach, it is different for each player i depending on the history player i has encountered up to time $t-1$. Since each group's adaptive learning path ($a^0(t)$) depends on its within-group interaction and $a^0(t)$ serves as the starting point of rule hierarchy, we adopted the closed-loop approach to better account for the level-0 rule dynamics.

[22] Note that we empirically assume that players make errors when they best respond, i.e., their actual choices are normally distributed around the predicted best-response. The standard deviation of these errors is empirically assumed to be σ, the same as that of $f(x, \mu)$, for the sake of parsimony. They could be assumed to be different in general.

Note that $\omega(0)$ and $\mathbf{B}(0)$ were separately estimated using the first round data. In sum, we estimated a total of 6 parameters $\Theta = (\alpha, \delta, \sigma_0, N(0), \rho, \sigma)$: for level-0 players, we estimated the probability that a player is only a level-0, adaptive-learning player throughout the game (α), the decaying rate of past winning actions in imitation learning (δ) and the error term for these adaptive level-0 players (σ_0); for Bayesian belief updating, we estimated the stickiness parameter of initial priors ($N(0)$), memory decay factor (ρ) and the error term for the sophisticated, belief-updating players (σ).

2.4 Estimated Models

We estimated a total of 4 models, which can be classified into two groups. The first group includes 2 models that incorporate adaptive level-0 players (i.e., $a^0(t)$) while the second group includes 2 models that do not incorporate adaptive level-0 players (i.e., $a^0(t) = a^0(1),\ \forall t$). The full model $BL_k(t)$ nests the other 3 models as special cases. By comparing the fit of our model with that of nested cases, we can determine the importance of adaptive and sophisticated learning. The first group includes the following 2 models:

1. Bayesian level-k model with adaptive level-0 ($BL_k(t)$): This is the full model with the parametric space $(\alpha, \delta, \sigma_0, N(0), \rho, \sigma)$ having 6 parameters. The first three parameters capture level-0's adaptive learning while the last three parameters capture higher level thinkers' sophisticated learning.
2. Static level-k model with adaptive level-0 ($L_k(t)$): This model allows level-0 players to adaptively learn and $a^0(t)$ to change over time. However, the rule levels of higher level thinkers remain fixed over time. This model is a nested case of $BL_k(t)$ with $N(0) = \infty$ and $\rho = 1$, i.e., the parametric space is $(\alpha, \delta, \sigma_0, \infty, 1, \sigma)$. If we further restrict $\alpha = 1$, $L_k(t)$ reduces to pure adaptive learning models which include experience-weighted attraction

learning, fictitious play (Brown, 1951), and reinforcement learning models.

The second group includes the following 2 models:

3. Bayesian level-k model with stationary level-0 (BL_k): This model ignores level-0's adaptive learning and only captures sophisticated learning. Specifically, it allows higher level thinkers to update their beliefs after each round about what rules others are likely to choose. As a result, these higher level thinkers choose a rule with different probabilities over time. This model is obtained by setting $\delta = 1$ in $BL_k(t)$, i.e., $a^0(t) = a^0(1)$, $\forall t$. The resulting parametric space is $(\alpha, 1, \sigma_0, N(0), \rho, \sigma)$.
4. Static level-k with stationary level-0 (L_k): This captures neither adaptive learning nor sophisticated learning. It is obtained by setting $N(0) = \infty$ and $\rho = \delta = 1$. The resulting parametric space is $(\alpha, 1, \sigma_0, \infty, 1, \sigma)$.

2.5 Estimation Results

Tables 1 and 2 show the estimation results. In both tables, the first column lists the names of parameters and test statistics. The 4 columns to the right of that report results from each of the four estimated models. The top panel reports parameter estimates and the bottom panel reports maximized log-likelihood and test statistics. We report χ^2, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).

The bottom panels show that $BL_k(t)$ fits the data best. All three nested cases are strongly rejected in favor of $BL_k(t)$. This is true regardless of which test statistic is used. Below we discuss the estimation results in detail.

2.5.1 p-beauty Contest Games

The bottom panel of Table 1 shows that $BL_k(t)$ fits the data much better than the three special cases. In fact, all three test statistics (χ^2, AIC, and

BIC) indicate that the three special cases are strongly rejected in favor of $BL_k(t)$. For instance, the χ^2 test statistics of L_k, BL_k, and $L_k(t)$ against $BL_k(t)$ are 3082.15, 744.52, and 336.58, respectively. These results suggest that allowing Bayesian updating by higher level thinkers and having adaptive level-0 players are both crucial in explaining subjects' dynamic choices over time. Put differently, higher level thinkers not only are aware that level-0 players adapt their choices, but also adjust their rule levels over time.[23]

[INSERT TABLE 1 HERE.]

The estimated parameters of $BL_k(t)$ appear reasonable. The fraction of adaptive level-0 players ($\hat{\alpha}$) is estimated to be 38%. These adaptive level-0 players assign equal weight to the most recent observed winning action and their own past choices (i.e., $\hat{\delta} = 50\%$), suggesting that players notice that the winning actions are converging towards equilibrium. The level-0 players' choices have a standard deviation of $\hat{\sigma_0} = 1.68$, and their choices are more variable than those of higher level thinkers ($\hat{\sigma} = 0.32$).[24]

The parameter estimates of higher level thinkers ($k \geq 1$) are also sensible. These higher level thinkers give their initial priors the weight $\hat{N}(0) = 3.35$ and decay this prior weight at a rate of $\hat{\rho} = 0.4$. That is, the influence of the initial prior drops to $3.35 \cdot 0.4^t$ after t rounds. Note that the standard deviation of level-k thinkers is estimated to be $\hat{\sigma} = 0.32$, implying that their choices are significantly less noisy than those of level-0 players.

We note the following patterns in other parameter estimates:

1. Comparing the left and right panels of Table 1, we observe that the choices of level-0 players become significantly less variable if they are allowed to

[23] We also fit a model of pure adaptive learning by setting $\alpha = 1$ (i.e., all players are adaptive level-0 players who engage in imitation learning). The estimated model has a maximized log-likelihood of -10341 and is strongly rejected in favor of $BL_k(t)$ with $\chi^2 = 1298$.

[24] In the estimation, we normalized subjects' choices by the sample standard deviation. As a consequence, a σ value of 1 means that choices are as variable as those made by an average player in the population.

adapt ($\hat{\sigma_0}$ drops from 19.990 to 1.565).

2. Allowing for either adaptive level-0 players or Bayesian updating of rule levels by higher level thinkers makes the latter players' choices less variable ($\hat{\sigma}$ drops from 1.10 to 0.320).

3. On the right panel of the table, we see that the fraction of level-0 players drops from 48% to 38%. This means that if we did not allow players to change their rule levels, the fraction of higher level thinkers would be underestimated.

4. Across the four models, allowing higher level thinkers to update their beliefs reduces the weight placed on the history ($\hat{\rho}$ drops from 1 to 0.345).

Figure 6a presents three-dimensional bar graphs of actual choice frequency of experimental data. Figures 6b-6e show the absolute errors (= |Predicted Choice Frequency − Actual Choice Frequency|) over time for the four models. In these absolute error plots, a perfect model will generate a bar graph of zero frequency in every choice bracket in every round. The poorer a model is, the higher its bar graph will be. Figure 7 shows median choices from actual data as well as those predicted by the four models.[25] Note the following:

1. We start with the simplest static level-k model, L_k. As shown in Figure 6e, this model fits the data poorly. In fact, its absolute error increases over time. Figure 7 confirms this and shows that L_k consistently over-predicts median choices from round 3 onward. This result is not surprising given that L_k ignores both adaptive and sophisticated learning. In this regard, our model represents a significant extension of L_k in terms of both theory and empirical fit.

[25]Note that both the actual and predicted data distributions become skewed towards the equilibrium as convergence takes place and as a result, their average choices become sensitive to off-equilibrium choices over time. Thus, we used the median instead of average choice for model comparisons.

2. Model BL_k allows higher-level thinkers to learn while keeping level-0 players fixed over time. As shown in Figure 6d, this model fits the data better than L_k, suggesting that sophisticated learning helps to capture choice dynamics. Figure 7 shows that after round 3, BL_k's predicted median choice is consistently closer to the actual median choice than L_k's. However, BL_k still consistently over-predicts actual choices (i.e., its median is always higher than the actual median).
3. Model $L_k(t)$ incorporates adaptive level-0 players by allowing them to imitate past winners. However, higher level thinkers are assumed to hold stationary beliefs about their opponents' rule levels. Figure 6c shows a significant reduction in bar heights compared to Figures 6d and 6e, suggesting that $L_k(t)$ yields predictions closer to actual data than BL_k and L_k. Figure 7 shows that $L_k(t)$ fits the data better than the other two nested models by significantly reducing the gap between predicted and actual median choices, starting in round 4. These patterns are also consistent with the result in Table 1 that $L_k(t)$ is the second best-fitting model.
4. Model $BL_k(t)$ allows for both adaptive level-0 players and higher level thinkers' sophisticated learning. Not surprisingly, Figure 6b shows that $BL_k(t)$ produces the lowest error bars among the four models. Compared to $L_k(t)$, the superiority of $BL_k(t)$ occurs in the earlier rounds (rounds 2 to 5). This suggests that adaptive learning alone fails to adequately capture actual choice dynamics in early rounds and that sophisticated learning helps close this gap. This superiority in earlier rounds is also demonstrated in Figure 7 where $BL_k(t)$ tracks the actual median choice better than $L_k(t)$ up to round 7, during which significant convergence to equilibrium occurs.

[INSERT FIGURE 6a-6e and 7 HERE.]

2.5.2 Generalized Price Matching Games

The bottom panel of Table 2 shows that $BL_k(t)$ fits the data best. All three test statistics (χ^2, AIC, and BIC) suggest that the three nested models are strongly rejected in favor of the $BL_k(t)$. For instance, the BIC test statistics were 14481.81, 12982.87, 12859.85 and 12699.05 for L_k, BL_k, $L_k(t)$ and $BL_k(t)$, respectively, in favor of $BL_k(t)$. These results suggest that both types of learning are pivotal in explaining choice dynamics in data.[26]

[INSERT TABLE 2 HERE.]

The estimated fraction of adaptive level-0 players ($\hat{\alpha}$) is 26%. These adaptive level-0 players assign a weight of $\hat{\delta} = 66\%$ to the past level-0's choice. As before, the choices of level-0 players also have a higher standard deviation than those of higher level thinkers ($\hat{\sigma_0} = 2.21$ versus $\hat{\sigma} = 0.03$).

Three quarters of the subjects are higher level thinkers. These players give their initial priors the weight $\hat{N}(0) = 10.23$ and decay this prior weight at a rate of $\hat{\rho} = 0.80$ after each round. That is, the influence of initial prior drops to $10.23 \cdot 0.80^t$ after t rounds.

Note the following on the parameter estimates reported in Table 2:

1. Comparing the left and right panels of Table 1, we see that level-0 players' choices become significantly less variable if they are allowed to adapt ($\hat{\sigma_0}$ drops from 19.990 to 1.985).
2. Allowing for either adaptive level-0 players or Bayesian belief updating introduces some variability into higher level thinkers' choices ($\hat{\sigma}$ increases from 0.00 to 0.027).
3. The right panel of the table shows that the fraction of level-0 players drops from 33% to 26%. This means if we did not allow higher level thinkers to so-

[26] We also fit a model of pure adaptive learning by setting $\alpha = 1$ (i.e., all players are adaptive, imitation learners). The estimated model has a maximized log-likelihood -6987.49 and is rejected in favor of $BL_k(t)$ with $\chi^2 = 1319$.

phisticatedly learn, the fraction of these players would be under-estimated.

4. Allowing higher level thinkers to update their beliefs reduces the weight placed on the history ($\hat{\rho}$ drops from 1 to 0.745).

In sum, these parameter estimates are remarkably similar to those from the p-beauty contest games. This is surprising, but encouraging, given that the two games are different and that subjects' choices converge towards equilibrium much faster in generalized price matching games.

[INSERT FIGURE 8a-8e and 9 HERE.]

Figure 8a presents three-dimensional bar graphs of actual choice frequency of experimental data. Figures 8b-8e show the absolute errors of the four models. Figure 9 compares median choices of actual data with those predicted by the four models. Note the following:

1. We again start with the static level-k model, L_k. As shown in Figure 8e, L_k shows the poorest fit. In fact, the absolute error of the model increases significantly over time, especially near the equilibrium as the data converges to it. Figure 9 confirms this result and shows that L_k's predicted median choice is static near 150 while the data shows clear convergence to 80. Hence, L_k consistently over-predicts choices in all rounds. This result is not surprising given that the model takes into account neither type of learning.
2. Model BL_k allows for higher level thinkers to learn about opponents' rule levels. Figure 8d shows that BL_k induces a significant reduction in the bar heights compared to L_k, suggesting that sophisticated learning helps to capture dynamics in this game. Figure 9 confirms this finding and shows that BL_k's predicted median is much closer to the actual median than L_k's after round 3. However, BL_k still shows significantly slower convergence than the actual data (i.e., its median becomes identical to the equilibrium solution only after round 7).

3. Model $L_k(t)$ incorporates adaptive level-0 players while not allowing for belief updating by higher level thinkers. Figure 8c shows a significant reduction in bar heights compared to Figure 8e (L_k), suggesting that $L_k(t)$ clearly dominates L_k in model prediction. Comparing Figures 8c and 8d, we notice that $L_k(t)$ fits higher choices better and lower choices worse. Since the data converges to a lower choice (80), misfitting lower choices hurts less than misfitting higher choices. This main result is consistent with $L_k(t)$'s better log-likelihood fit than BL_k's, shown in Table 2; and also with $L_k(t)$'s closer median curve to the actual median than BL_k's, shown in Figure 9.
4. $BL_k(t)$, the full model, allows for both adaptive and sophisticated learning. As shown in Figure 8b, $BL_k(t)$ produces the lowest error bars. It is worth highlighting that compared to the second best-fitting $L_k(t)$, most of the error reduction comes in earlier rounds (rounds 2 to 5), which is consistent with the pattern observed in the p-beauty contest games. This again suggests a significant role of sophisticated learning in explaining choice dynamics before the adaptive learning takes on a full speed (i.e., before players notice that choices clearly converge to equilibrium). Consistent with this result is Figure 9. In Figure 9, the median of $BL_k(t)$ becomes identical to the actual median in round 4 whereas it happens for $L_k(t)$ in round 6.

3 Conclusion

The level-k model is an attractive candidate to explain nonequilibrium behavior because it is based on a highly intuitive iterative reasoning process. Each iteration in the level-k model is akin to a thinking step. Players start at level-0 and iteratively advance one step by best-responding to their counterparts who reason one step less. Nonequilibrium behavior emerges when players do not perform enough thinking steps. While this level-k modeling approach has been successfully applied to both laboratory and field experiments, it is far from complete since the level-k model is inherently static. A wide body of empirical evidence shows that behavior converges to equilibrium over time.

To capture such dynamic choice behavior, this paper proposes a Bayesian level-k model that distinguishes between two possible types of learning dynamics. One possibility is that the starting point of the iterative reasoning process (i.e., level-0) moves towards equilibrium, so players approach the equilibrium while performing the same number of thinking steps throughout. Another possibility is that players approach the equilibrium by engaging in more thinking steps. In sum, our model generalizes the standard level-k model in two significant ways:

1. We allow the level-0 rule to adaptively respond to historical game plays (e.g., past winning or chosen actions). To the extent that observed game plays move towards equilibrium, players may approach the equilibrium without advancing in rule levels. In this way, players exhibit adaptive learning.
2. We allow players to revise their rule levels by Bayesian updating, where players update beliefs about opponents' rule levels based on observations. Thus, players perform more thinking steps when they anticipate likewise from opponents. In this way, players exhibit sophisticated learning.

Interestingly, the extended model provides a novel unification of two separate streams of nonequilibrium models: level-k and adaptive learning models. On one hand, level-k models are "static" and have been used to predict behavior in one-shot games. On the other hand, adaptive learning models have been used to describe choice dynamics in repeated games. While these two seemingly distinct classes of models have been treated separately in the literature, our model brings them together and provides a sensible way to model behavior in both one-shot and repeated games in a common, tractable framework.

We apply the Bayesian level-k model to experimental data on two games: p-beauty contest (from Ho et al. (1998)) and generalized price matching (freshly collected through new experiments). For both games, we find strong evidence that our model explains the data better than all its special cases and that it is essential to account for sophisticated learning even after existing nonequilibrium models are introduced. In other words, beyond merely best responding to historical game plays, players dynamically adjust their rule levels over time in response to changing anticipation about opponents' rule levels.

By incorporating both adaptive and sophisticated learning, our model can be viewed as a characterization of the equilibration process. This view bears the same spirit as Harsanyi's "tracing procedure" (Harsanyi and Selten, 1988), in which players' successive choices, as they react to new information in strategic environments, trace a path towards the eventual equilibrium outcome.

References

Basu, K. 1994. "The Traveler's Dilemma: Paradoxes of Rationality in Game Theory." Am. Econ. Rev. 84(2): 391-395.

Binmore, K., A. Shaked and J. Sutton. 1985. "Testing Noncooperative Bargaining Theory: A Preliminary Study." Am. Econ. Rev. 75: 1178-1180.

Bosch-Doménech, A., J. G. Montalvo, R. Nagel and A. Satorra. 2002. "One, Two, (Three), Infinity, ... : Newspaper and Lab Beauty-Contest Experiments." Am. Econ. Rev. 92(5): 1687-1701.

Bosch-Doénech, A. and N. J. Vriend. 2003. "Imitation of successful behaviour in the markets." The Economic Journal 113: 495524.

Brown, A., C. F. Camerer and D. Lovallo. 2012. "To Review or Not To Review? Limited Strategic Thinking at the Movie Box Office." American Economic Journal: Microeconomics 4(2): 1-28.

Brown, George W. 1951. "Iterative solution of games by fictitious play." Activity Analysis of Production and Allocation 13(1): 374-376.

Camerer, C. and T-H. Ho. 1999. "Experience-Weighted Attraction Learning in Normal Form Games." Econometrica 67: 837-874.

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

Chen, Y., C. Narasimhan and Z. J. Zhang. 2001. "Consumer heterogeneity and competitive price-matching guarantees." Market Sci. 20(3): 300-314.

Capra, C. M., J. K. Goeree, R. Gomez and C. A. Holt. 1999. "Anomalous Behavior in a Traveler's Dilemma?." Am. Econ. Rev. 89(3): 678-690.

Costa-Gomes, M., V. Crawford and B. Broseta. 2001. "Cognition and Behavior in Normal-Form Games: An Experimental Study." Econometrica 69: 1193-1235.

Costa-Gomes, M. and V. Crawford. 2006. "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study." Am. Econ. Rev. 96: 1737-1768.

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

Costa-Gomes, M., V. Crawford and N. Iriberri. 2009. "Comparing Models of Strategic Thinking in Van Huyck, Battalio, and Beil's Coordination Games." Journal of the European Economic Association 7: 377-387.

Crawford, V. P. 2003. "Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intentions." Am. Econ. Rev. 93: 133-149.

Crawford, V. P. and N. Iriberri. 2007a. "Fatal Attraction: Salience, Naivete, and Sophistication in Experimental Hide-and-Seek Games." Am. Econ. Rev. 97: 1731-1750.

Crawford, V. P. and N. Iriberri. 2007b. "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.

Crawford, V. P., T. Kugler, Z. Neeman and A. Pauzner. 2009. "Behaviorally Optimal Auction Design: An Example and Some Observations." Journal of the European Economic Association 7(2-3): 377-387.

Crawford, V. P., M. A. Costa-Gomes and N. Iriberri. 2013. "Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications." Journal of Economic Literature, In Press.

Ellison, G. and D. Fudenberg. 2000. "Learning Purified Mixed Equilibria." J. Econ. Theory 90(1): 84-115.

Fudenberg, D. and D. Kreps. 1993. "Learning mixed equilibria." Games and Economic Behavior 5: 320-367

Fudenberg, D. and D. Levine. 1993. "Steady State Learning and Nash Equilibrium." Econometrica 61(3): 547-573.

Fudenberg, D. and D. Levine. 1998. "Theory of Learning in Games." Cambridge, MIT Press.

Goldfarb, A. and B. Yang. 2009. "Are All Managers Created Equal?." J. Marketing Res. 46(5): 612-622.

Goldfarb, A. and M. Xiao. 2011. "Who Thinks About the Competition? Managerial Ability and Strategic Entry in US Local Telephone Markets." Am. Econ. Rev. 101(7): 3130-3161.

Goeree, J. K. and C. A. Holt. 2001. "Ten Little Treasures of Game Theory and Ten Intuitive Contradictions." Am. Econ. Rev. 91(5):1402-1422.

Harsanyi, J. C. and R. Selten. 1988. "A General Theory of Equilibrium Selection in Games." Cambridge, MIT Press.

Hück, S., H. Normann and J. Oechssler. 1999. "Learning in Cournot Oligopoly–An Experiment." Economic Journal 109: C80-C95.

Ho, T-H., C. Camerer and K. Weigelt. 1998. "Iterated Dominance and Iterated Best Response in Experimental p-Beauty Contests." Am. Econ. Rev. 88: 947-969.

Ho, T-H., C. Camerer, C. and J-K. Chong. 2007. "Self-tuning Experience-Weighted Attraction Learning in Games." J. Econ. Theory 133(1): 177-198.

Ho, T-H. and X. Su. 2013. "A Dynamic Level-k Model in Sequential Games." Manage. Sci. 59(2): 452-469.

Jain, S. and J. Srivastava. 2000. "An experimental and theoretical analysis of price-matching refund policies." J. Marketing Res. 37(3): 351-362.

Nagel, R. 1995. "Unraveling in Guessing Games: A Experimental Study." Am. Econ. Rev. 85(5): 1313-26.

McKelvey, R. D. and T. R. Palfrey. 1992. "An Experimental Study of the Centipede Game." Econometrica 60: 803-836.

Ochs, J. and A. E. Roth. 1989. "An Experimental Study of Sequential Bargaining." Am. Econ. Rev. 79: 355-384.

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

Srivastava, J. and N. Lurie. 2001. "A Consumer Perspective on Price? Matching Refund Policies: Effect on Price Perceptions and Search Behavior." Journal of Consumer Research 28(2): 296-307.

Stahl, D. and P. Wilson. 1994. "Experimental Evidence on Players' Models of Other Players." J. Econ. Behav. Organ. 25 (3): 309-327.

Stahl, D. and P. Wilson. 1995. "On Players' Models of Other Players - Theory and Experimental Evidence." Games and Economic Behavior 10: 213-254.

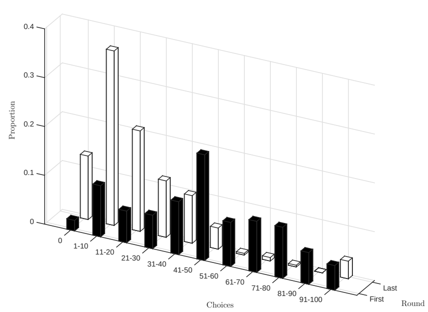

Figure 1: p-beauty Contest: First and Last Round Data

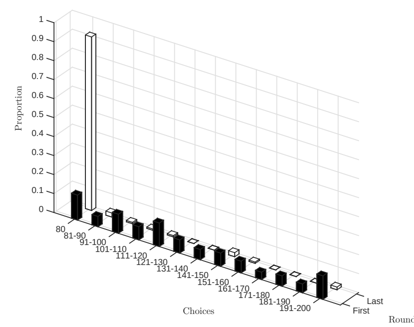

Figure 2: Generalized Price Matching: First and Last Round Data

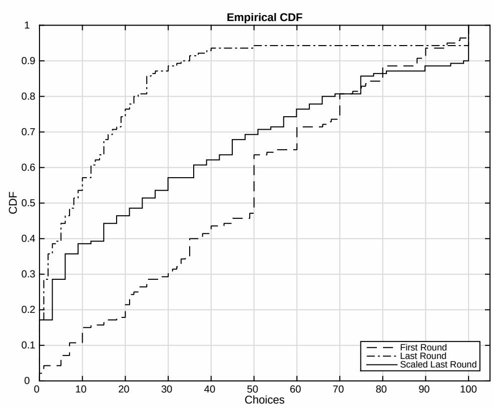

Figure 3: p-beauty Contest: Distributions of First Round, Last Round, and Scaled Last Round Data

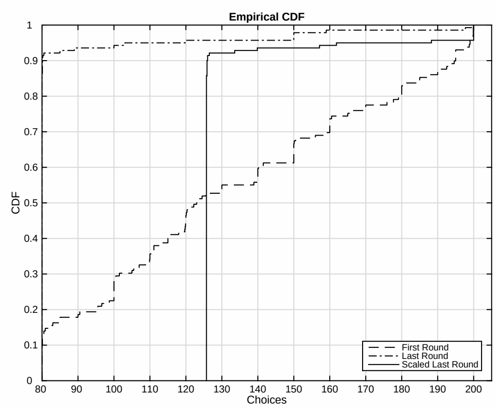

Figure 4: Generalized Price Matching: Distributions of First Round, Last Round, and Scaled Last Round Data

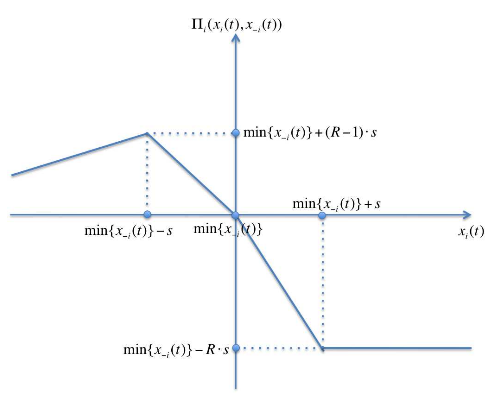

Figure 5: Payoff Function of Generalized Price Matching Game

Table 1: p-beauty Contest: Parameter Estimates

	Models			
	Static $a_i^0(t)$		Dynamic $a_i^0(t)$	
	Static Level-k	Bayesian	Static Level-k	Bayesian
	L_k	BL_k	$L_k(t)$	$BL_k(t)$
Parameters[a]				
Level-0 Players				
$\hat{\alpha}$	0.04	0.32	0.48	0.38
	(0.03, 0.08)	(0.21, 0.34)	(0.46, 0.49)	(0.36, 0.39)
$\hat{\delta}$	$\underline{0}$	$\underline{0}$	0.19	0.50
			(0.17, 0.20)	(0.44, 0.51)
$\hat{\sigma_0}$	20.00	19.98	1.45	1.68
	(20.00, 20.00)	(15.89, 20.00)	(1.42, 1.47)	(1.62, 1.72)
Bayesian Updating				
$\hat{N_0}$	$\underline{\infty}$	2.81	$\underline{\infty}$	3.35
		(2.68, 3.11)		(3.29, 3.42)
$\hat{\rho}$	$\underline{1}$	0.29	$\underline{1}$	0.40
		(0.28, 0.36)		(0.39, 0.41)
$\hat{\sigma}$	1.10	0.33	0.31	0.32
	(1.07, 1.12)	(0.29, 0.63)	(0.29, 0.31)	(0.32, 0.33)
LL	-11233.98	-10065.17	-9861.20	-9692.91
χ^2	3082.15	744.52	336.58	–
(p-value, dof)	(0.000, 3)	(0.000, 1)	(0.000, 2)	–
AIC[b]	22473.96	20140.33	19730.40	19397.82
BIC[c]	22491.43	20169.44	19753.68	19432.75

[a]Restricted parameter values in nested models are marked with the underline. Results from burning the first round data are: $\hat{\omega}(0) = 68.72$ and $\hat{\hat{\sigma}} = 0.89$; prior belief distribution is estimated using 50 rules in the rule hierarchy and top two rules with highest probability mass were $\hat{B}^0(0) = 0.50$ and $\hat{B}^2(0) = 0.27$. Maximized log-likelihood is $LL = -357.27$.

[b]$\text{AIC} = 2 \cdot k - 2 \cdot LL$ where k is the number of free parameters.

[c]$\text{BIC} = k \cdot \ln(N) - 2 \cdot LL$ where k is the number of free parameters and N is the number of data points.

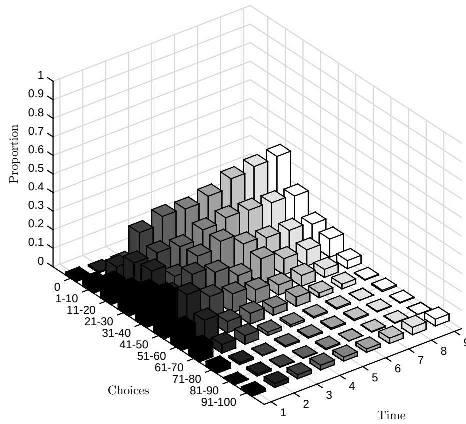

Figure 6a: p-Beauty Contest: Actual Data

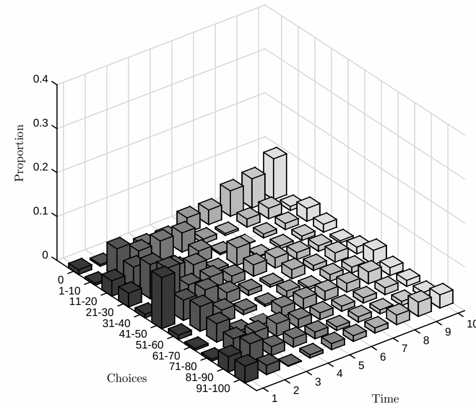

Figure 6b: p-beauty Contest: Absolute Errors of $BL_k(t)$

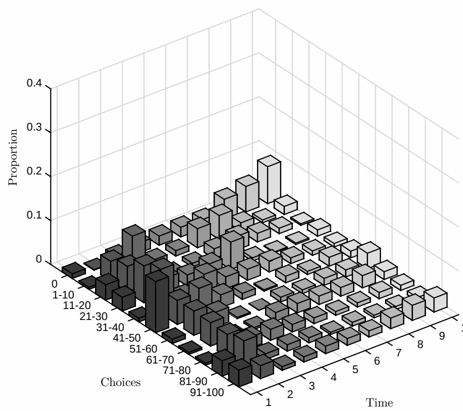

Figure 6c: p-beauty Contest: Absolute Errors of $L_k(t)$

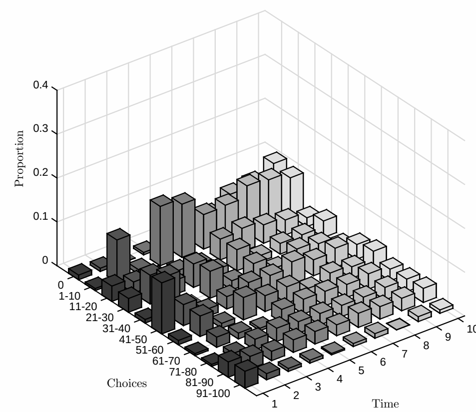

Figure 6d: p-beauty Contest: Absolute Errors of BL_k

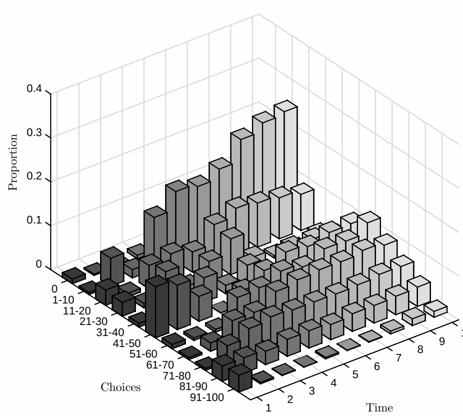

Figure 6e: p-beauty Contest: Absolute Errors of L_k

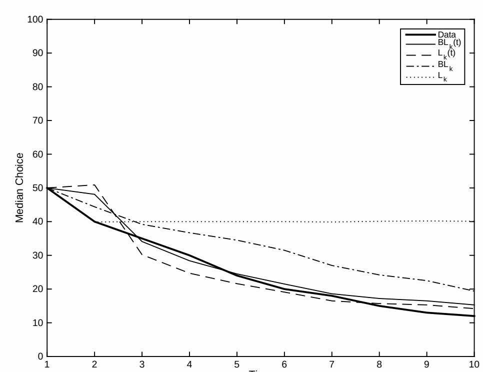

Figure 7: p-beauty Contest: Median of Actual and Predicted Data

Table 2: Generalized Price Matching: Parameter Estimates

	Models			
	Static $a_i^0(t)$		Dynamic $a_i^0(t)$	
	Static Level-k	Bayesian	Static Level-k	Bayesian
	L_k	BL_k	$L_k(t)$	$BL_k(t)$
Parameters [a]				
Level-0 Players				
$\hat{\alpha}$	0.59	0.25	0.33	0.26
	(0.19, 0.61)	(0.17, 0.26)	(0.25, 0.36)	(0.23, 0.33)
$\hat{\delta}$	$\underline{0}$	$\underline{0}$	0.49	0.66
			(0.37, 0.66)	(0.55, 0.73)
$\hat{\sigma_0}$	19.98	20.00	1.76	2.21
	(19.92, 20.00)	(19.06, 20.00)	(1.66, 2.30)	(1.82, 2.79)
Bayesian Updating				
$\hat{N}_0$	$\underline{\infty}$	6.51	$\underline{\infty}$	10.23
		(6.16, 7.32)		(9.35, 11.87)
$\hat{\rho}$	$\underline{1}$	0.69	$\underline{1}$	0.80
		(0.68, 0.73)		(0.79, 0.83)
$\hat{\sigma}$	0.00	0.02	0.03	0.03
	(0.00, 0.03)	(0.02, 0.03)	(0.03, 0.03)	(0.03, 0.03)
LL	-7230.03	-6473.30	-6415.42	-6327.76
χ^2	1804.52	291.07	175.31	–
(p-value, dof)	(0.000, 3)	(0.000, 1)	(0.000, 2)	–
AIC[b]	14466.05	12956.60	12838.84	12667.53
BIC[c]	14481.81	12982.87	12859.85	12699.05

[a]Restricted parameter values in nested models are marked with the underline. Results from burning the first round data are: $\hat{\omega}(0) = 200.00$ and $\hat{\tilde{\sigma}} = 1.58$; prior belief distribution is estimated using 25 rules in the rule hierarchy and top two rules with highest probability mass were $\hat{B}^0(0) = 0.50$ and $\hat{B}^{10}(0) = 0.08$. Maximized log-likelihood is $LL = -1025.21$.

[b]$\text{AIC} = 2 \cdot k - 2 \cdot LL$ where k is the number of free parameters.

[c]$\text{BIC} = k \cdot \ln(N) - 2 \cdot LL$ where k is the number of free parameters and N is the number of data points.

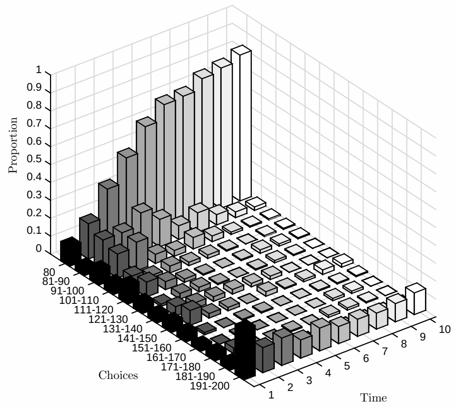

Figure 8a: Generalized Price Matching: Actual Data

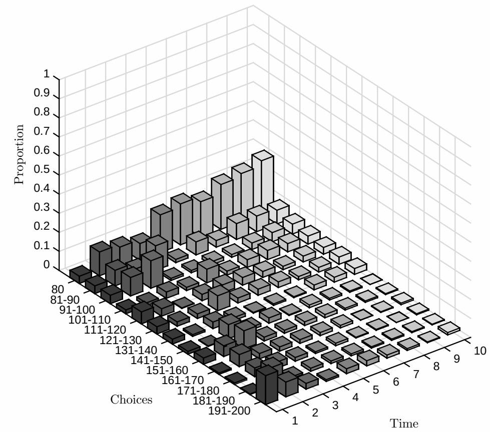

Figure 8b: Generalized Price Matching: Absolute Errors of $BL_k(t)$

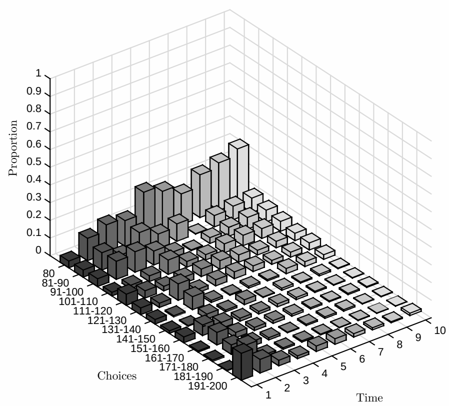

Figure 8c: Generalized Price Matching: Absolute Errors of $L_k(t)$

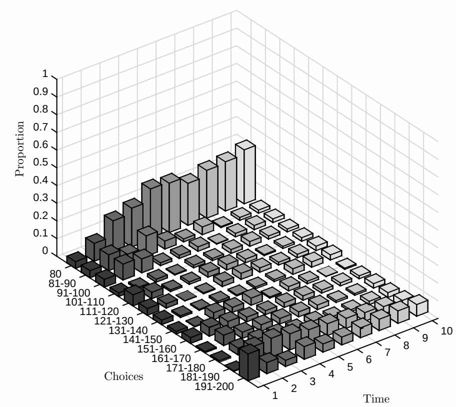

Figure 8d: Generalized Price Matching: Absolute Errors of BL_k

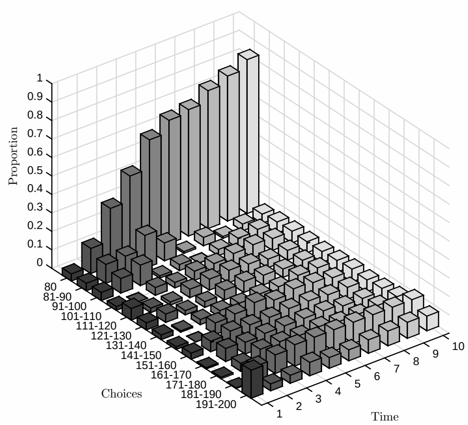

Figure 8e: Generalized Price Matching: Absolute Errors of L_k

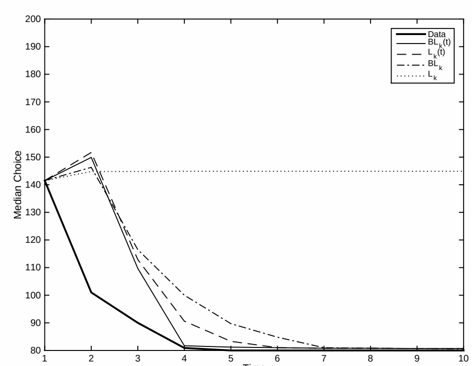

Figure 9: Generalized Price Matching: Median of Actual and Predicted Data

Online Appendix

A. Proof of Proposition 1

(1) if $a_i^0(t) \to x_i^*$, then by the definition of rule hierarchy, $a_i^k(t) \to x_i^*$ $\forall k \geq 1$. If player i is level-0, then $x_i(t) = a_i^0(t) \to x_i^*$ as $t \to \infty$. If player i is higher level, $\mathbb{E}[x_i(t)] = \sum_{k=0}^{\infty} B_i^k(t-1) \cdot a_i^k(t) \to x_i^*$ as $t \to \infty$.

(2) If $a_i^0(t) \not\to x_i^*$ $\exists\, i$ and all conditions (a)-(d) are met, we prove by induction. Without loss of generality, suppose $a_i^0(t) = a_i^0(0)$, $\forall t, \forall i$. (Proof is analogous if $|a_i^0(t) - x_i^*| > \delta > 0$ for all t big enough for some i.) Since $\alpha = 0$ by condition (a), it suffices to show that $B_i^k(t) \to 0$ as $t \to \infty$, $\forall k \geq 0$, $\forall i$. Let us first prove the base case $k = 0$. Note that:

$$B_i^0(t) = \frac{\rho^t N_i(0) B_i^0(0) + (n-1)\sum_{\tau=0}^{t-1} \rho^\tau P_{-i}^0(\tau+1)}{\rho^t N_i(0) + (n-1)\sum_{\tau=0}^{t-1} \rho^\tau}$$

Thus, $B_i^0(t)$ is the weighted sum of $B_i^0(0)$, $P_{-i}^0(1)$, $P_{-i}^0(2)$,...., $P_{-i}^0(t)$ (or in other words, the weighted sum of $B_i^0(t-1)$ and $P_{-i}^0(t)$) and the weight given to history bears less influence over time since $N_i(0) < \infty$ by condition (b). Hence, it suffices to show that $P_{-i}^0(t) \to 0$ as $t \to \infty$. In order to show this, we first show that $P_{-i}^0(t) < B_i^0(t-1) \cdot \gamma$, $\exists \gamma \in (0,1)$, where γ is a constant, and $B_i^0(t) < B_i^0(t-1)$, $\forall t, \forall i$. We prove these two statements by induction. Suppose they hold for time up to $t-1$. (Proof of the base case $t = 1$ is exactly analogous to proof of the inductive step below.) Then, by definition,

$$P_{-i}^0(t) = \frac{B_i^0(t-1) \cdot \prod_{j \neq i} f(a_j^{k_j(t)}(t), a_j^0(t))}{\sum_{k=0}^{\infty} B_i^k(t-1) \cdot \prod_{j \neq i} f(a_j^{k_j(t)}(t), a_j^k(t))}$$

Since all players are higher level thinkers, $k_j(t) \geq 1$ for all j and t. Thus, by condition (d), for each realization of $\{k_j(t) \geq 1 \,|\, j \neq i\}$, we have: $1 < \frac{\prod_{j \neq i} f(a_j^{k_j(t)}(t), a_j^k(t))}{\prod_{j \neq i} f(a_j^{k_j(t)}(t), a_j^0(t))}$ if $k \in \{k_j(t)\}$ and $1 = \frac{\prod_{j \neq i} f(a_j^{k_j(t)}(t), a_j^k(t))}{\prod_{j \neq i} f(a_j^{k_j(t)}(t), a_j^0(t))}$ if $k \notin \{k_j(t)\}$. Let

$m = \inf_{k,k'} \frac{f(a_i^k(t), a_i^k(t))}{f(a_i^k(t), a_i^{k'}(t))}$, where $m > 1$ by condition (d). Then, $P_{-i}^0(t) = B_i^0(t-1) \cdot \frac{1}{\sum_{k=0}^{\infty} B_i^k(t-1) \cdot \frac{\prod_{j \neq i} f(a_j^{k_j(t)}(t), a_j^k(t))}{\prod_{j \neq i} f(a_j^{k_j(t)}(t), a_j^0(t))}} < B_i^0(t-1) \cdot \frac{1}{(1 - \sum_{k \in \{k_j(t)\}} B_i^k(t-1)) + \sum_{k \in \{k_j(t)\}} B_i^k(t-1) \cdot m}$ $\leq B_i^0(t-1) \cdot \frac{1}{(1 - B_i^0(0)) + B_i^0(0) \cdot m}$ Thus, $P_{-i}^0(t) < B_i^0(t-1) \cdot \gamma$, where $\gamma = \frac{1}{(1 - B_i^0(0)) + B_i^0(0) \cdot m} < 1$. The first inequality holds by condition (c) that the support of initial prior is the entire rule hierarchy and thus, for each realization of $\{k_j(t) \geq 1 \,|\, j \neq i\}$, there exists k where $1 < \frac{\prod_{j \neq i} f(a_j^{k_j(t)}(t), a_j^k(t))}{\prod_{j \neq i} f(a_j^{k_j(t)}(t), a_j^0(t))}$ and $B_i^k(t-1) > 0$. The second inequality holds because if prior is uniform, then $B_i^0(0) = B_i^k(0)$, $\forall k$, hence $P_{-i}^0(1) \leq P_{-i}^k(1)$, $\forall k$ for any realizations of $\{k_j(1)\}$ (because of condition (d) and the fact that $k \geq 1$, $\forall k \in \{k_j(t)\}$), hence $B_i^0(1) \leq B_i^k(1)$, $\forall k$, and so forth. Thus, $P_{-i}^0(t) \leq P_{-i}^k(t)$, $\forall k, \forall t$, and we have $B_i^0(0) \leq B_i^k(t)$, $\forall k, \forall t$. Hence, $P_{-i}^0(t) < B_i^0(t-1) \cdot \gamma$ and $B_i^0(t) < B_i^0(t-1)$ since $B_i^0(t)$ is the weighted sum of $B_i^0(t-1)$ and $P_{-i}^0(t)$.

Now we prove $\lim_{t\to\infty} P_{-i}^0(t) \to 0$ given that $P_{-i}^0(t) < B_i^0(t-1) \cdot \gamma$, $\exists \gamma \in (0,1)$, where γ is a constant, and $B_i^0(t) < B_i^0(t-1)$, $\forall t, \forall i$. Since $B_i^0(t)$ is the weighted sum of $B_i^0(t-1)$ and $P_{-i}^0(t)$ by the definition given above, $\lim_{t\to\infty} B_i^0(t) \to (1-w) \cdot B_i^0(t-1) + w \cdot P_{-i}^0(t)$ where $w = 1 - \rho = \lim_{t\to\infty} \frac{n-1}{\rho^t N_i(0) + (n-1)\sum_{\tau=0}^{t-1} \rho^\tau} > 0$ by condition (b). Thus, there exists $\exists T$ such that $\forall t > T$, $B_i^0(t) < (1 - w + w \cdot \gamma) \cdot B_i^0(t-1)$. Hence, we have $P_{-i}^0(t+1) < B_i^0(t) \cdot \gamma < (1 - w + w \cdot \gamma) \cdot B_i^0(t-1) \cdot \gamma < (1 - w + w \cdot \gamma)^{t-T} \cdot B_i^0(T) \cdot \gamma$. Since $(1 - w + w \cdot \gamma) < 1$ ($\because w > 0$), we have $\lim_{t\to\infty} P_{-i}^0(t) \to 0$, $\forall i$ and $\lim_{t\to\infty} B_i^0(t) \to 0$, $\forall i$ as desired.

Now let us prove the inductive step. Suppose that $\forall \epsilon > 0$, there exists $\exists t_{k-1}(\epsilon)$ such that $\forall t > t_{k-1}(\epsilon)$, $B_i^{k'}(t) < \epsilon$, $\forall k' \leq k-1, \forall i$. Showing that there exists $\exists t_k(\epsilon) > t_{k-1}(\epsilon)$ such that $\forall t > t_k(\epsilon)$, $B_i^k(t) < \epsilon$, $\forall i$, will complete the proof. The proof is analogous to that of the base case and it suffices to show that

$\lim_{t\to\infty} P^k_{-i}(t) = 0$, $\forall i$. By definition, $P^k_{-i}(t) = \frac{B^k_i(t-1)\cdot\prod_{j\neq i} f(a_j^{k_j(t)}(t), a_j^k(t))}{\sum_{k'=0}^{\infty} B_i^{k'}(t-1)\cdot\prod_{j\neq i} f(a_j^{k_j(t)}(t), a_j^{k'}(t))}$

$= B^k_i(t-1) \cdot \frac{1}{\sum_{k'=0}^{\infty} B_i^{k'}(t-1)\cdot \frac{\prod_{j\neq i} f(a_j^{k_j(t)}(t), a_j^{k'}(t))}{\prod_{j\neq i} f(a_j^{k_j(t)}(t), a_j^{k}(t))}}$

$= B^k_i(t-1) \cdot \frac{1}{\sum_{k'=0}^{k-1} B_i^{k'}(t-1)\cdot \frac{\prod_{j\neq i} f(a_j^{k_j(t)}(t), a_j^{k'}(t))}{\prod_{j\neq i} f(a_j^{k_j(t)}(t), a_j^{k}(t))} + \sum_{k'=k}^{\infty} B_i^{k'}(t-1)\cdot \frac{\prod_{j\neq i} f(a_j^{k_j(t)}(t), a_j^{k'}(t))}{\prod_{j\neq i} f(a_j^{k_j(t)}(t), a_j^{k}(t))}}$

Note that $\forall k' \geq k+1$, $\frac{\prod_{j\neq i} f(a_j^{k_j(t)}(t), a_j^{k'}(t))}{\prod_{j\neq i} f(a_j^{k_j(t)}(t), a_j^{k}(t))} > m > 1$ by condition (d) if $k' \in \{k_j(t)\}$ and it equals 1 otherwise. Since $\sup_{k,k',\{k_j(t)\}} \frac{\prod_{j\neq i} f(a_j^{k_j(t)}(t), a_j^{k'}(t))}{\prod_{j\neq i} f(a_j^{k_j(t)}(t), a_j^{k}(t))} < \infty$ by condition (d), there exists ϵ' small enough where there exists $\exists t_{k-1}(\epsilon') > t_{k-1}(\epsilon)$ such that $\forall t > t_{k-1}(\epsilon')$, $P^k_{-i}(t) < B^k_i(t-1)\cdot\gamma(k)$, where $\gamma(k) = \frac{1}{(1-B_i^0(0)) + B_i^0(0)\cdot m} - \epsilon'$ so that $\gamma(k) > 1$. Hence, $P^k_{-i}(t) < B^k_i(t-1)\cdot\gamma(k)$ and $B^k_i(t) < B^k_i(t-1)$, $\forall i, \forall t > t_{k-1}(\epsilon')$, and following the steps analogous to the proof for the base case, we can show that $\lim_{t\to\infty} P^k_{-i}(t) \to 0$. Thus, there exists $\exists t_k(\epsilon) > t_{k-1}(\epsilon') > t_{k-1}(\epsilon)$ such that $\forall i, \forall t > t_k(\epsilon)$, $B^k_i(t) < \epsilon$ as desired. Q.E.D.

B. Instructions for Generalized Price Matching

Below is the instruction for the experimental generalized price matching game for the group of size 3. That for the group of size 7 is exactly same but only with different numerical examples.

Instructions

This is an experiment about economic decision-making. The instructions are simple; and if you follow them carefully and make good decisions, you may earn a considerable amount of money which will be paid to you in cash at the end of the experiment. It is important that you do not look at the decisions of others, and that you do not talk, laugh, or make noises during the experiment. You will be warned if you violate this rule the first time. If you violate this rule twice, you will be asked to leave the room immediately and your cash earnings will be $0.

The experiment consists of 10 decision rounds. In each decision round, you will be randomly assigned into groups of 3 participants. That is, you are randomly matched with other participants in each round, and your paired members may be different round by round.

Experimental Procedure

In each decision round, all members in each group are asked to choose a number simultaneously between 80 and 200 (inclusive of 80 and 200, and up to 1 decimal point). Specifically, each member can choose any number among 80.0, 80.1, 80.2,., 199.8, 199.9, and 200.0. Note that each member must make a choice without knowing what his or her paired members will choose. Once all members in a group make their choices, the computer will determine each members point earning based on everyones chosen number. Ones point earning is determined by ones chosen number and the minimum choice of ones paired members excluding oneself. Details of how point earning is determined will be given below. After each decision round, each participant will be told (a) his/her choice in that round, (b) the minimum choice of his/her paired members excluding him/herself in that round, and (c) his/her point earning in that round. Again, every participant will undertake this task 10 times, and will be randomly matched with other participants in each round.

Determination of Point Earnings

Your total point earning in each round is the sum of 2 components: baseline earning and supplementary earning. Specifically, we have:

TOTAL POINT EARNING = BASELINE EARNING + SUPPLEMENTARY EARNING.

1. BASELINE EARNING: In each round, the baseline earning in a group is given by the minimum of all members choices in that group. That is, we will examine all choices in each group and use the minimum of the choices in the group to determine its baseline earning. Note that all members in a group receive the same baseline earning. Different groups, however, may receive different baseline earnings depending on the choices of their respective group members.
2. SUPPLEMENTARY EARNING: Depending on the relative magnitude between your own choice and the minimum of your paired members choices, your supplementary earning can be either positive, negative, or zero. Each case is described in full detail below.

    (a) Positive Supplementary Earning: You will receive a positive supplementary earning (on top of your baseline earning) if your own choice is smaller than the minimum choice of your paired members excluding yourself. This means that your choice is the only smallest number in the group. That is, you have chosen the lowest number, while all your paired members have chosen bigger numbers than you did. Since you receive a positive supplementary earning, your total point earning will be higher than your baseline earning as a consequence.

        Precisely, your positive supplementary earning is 1.5 TIMES the difference between your choice and the minimum of your paired members choices. However, there is a cap on the maximum supplementary earning that you can earn. In this experiment, you can earn up to 15 supplementary points. Again, your total earning is the sum of your baseline earning and supplementary earning.

    (b) Negative Supplementary Earning: You will receive a negative supplementary earning (i.e., a deduction from your baseline earning) if your choice is higher than the minimum choice of your paired members excluding yourself. This means that someone else in your group has chosen a smaller number than you did, and your choice is NOT the lowest number (or the minimum). Since you receive a negative supplementary earning, your total point earning will be lower than your baseline earning as a consequence.

Precisely, the magnitude of your negative supplementary earning is 1.5 TIMES the difference between your choice and the minimum of your paired members choices. However, there is a cap on the maximum supplementary points that will be deducted from your baseline earning. In this experiment, up to 15 supplementary points can be deducted from your baseline earning. Again, your total earning is the sum of your baseline earning and supplementary earning.

(c) No Supplementary Earning: You will receive no supplementary earning if your choice is the smallest in your group, but there are also other paired members who have chosen the same number as well. As a consequence, your total point earning will simply be your baseline earning.

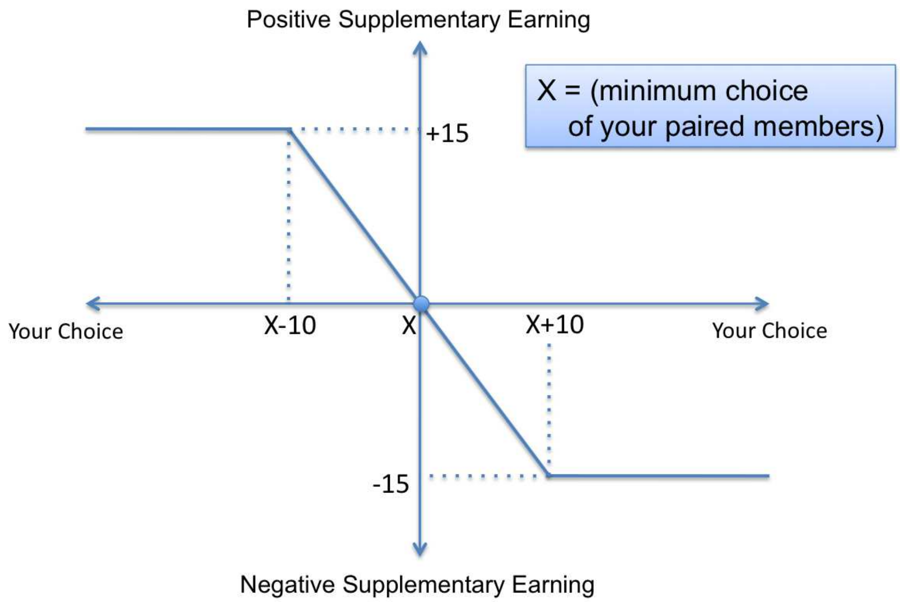

Figure 1: Supplementary Earning

Figure 1 summarizes the 3 possible cases of supplementary earning described above.

If your choice is smaller than the minimum of your paired members choices (on the left hand side of Figure 1), your positive supplementary earning increases as your choice gets even smaller and farther away from your paired members minimum choice (indicated as X at the origin). Your positive supplementary earning is exactly 1.5 times the difference between your choice and your paired members minimum choice. However, once the positive supplementary earning hits 15, your positive supplementary earning stops increasing and remains at 15. Note that as you choose a smaller number on the left hand side of Figure 1, the baseline earning for everyone in the group becomes smaller too (because the smaller number that you choose is also the minimum of all members choices).

If your choice is higher than the minimum of your paired members choices (on the right hand side of Figure 1), you will receive a negative supplementary earning. The magnitude of your negative supplementary earning increases as your choice gets even higher and farther away from the minimum of your paired members choices (indicated as X at the origin). The magnitude of your negative supplementary earning is exactly 1.5 times the difference between your choice and your paired members minimum choice. Likewise, once the negative supplementary earning hits -15, it stops getting larger and remains at -15. Note that as you choose a higher number on the right hand side of Figure 1, the baseline earning for everyone in the group remains the same at the minimum of your paired members choices (i.e., X) because X remains the minimum of all members choices.

Finally, if your choice is the same as the minimum of your paired members choices (when your choice is at the origin of Figure 1), your supplementary earning is zero, and your total earning is simply the baseline earning.

Illustrative Examples

The numbers used in following examples are chosen purely for illustrative purposes.

EXAMPLE 1: Assume your paired members chose 160.0 and 167.2, respectively. Note that the minimum of your paired members choices is 160.0 (the smaller of 160.0 and 167.2). We shall use Figure 2 below (which we created by relabeling Figure 1) to determine your supplementary earning depending on possible choices you can make. Note that the origin is now set to l60.

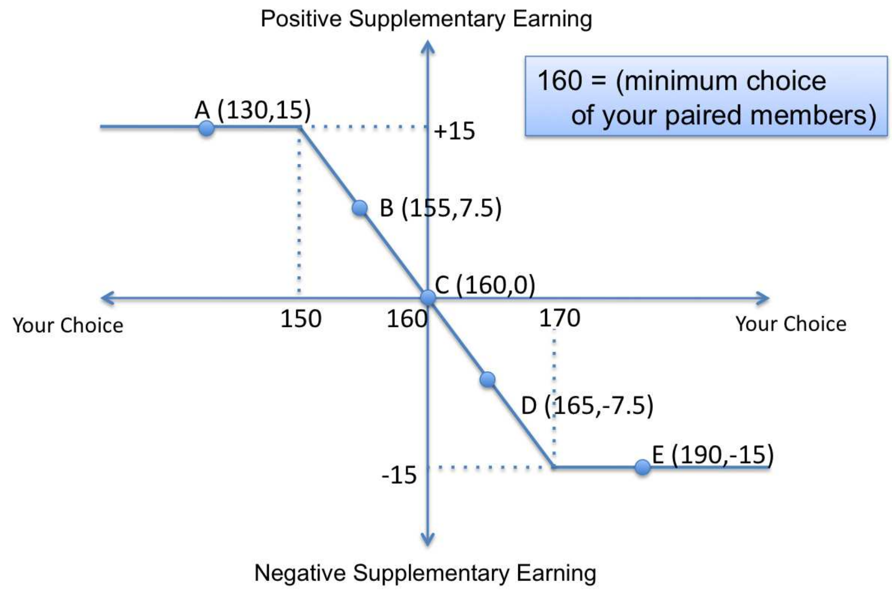

Figure 2: Choice Scenarios When the Paired Members Chose 160.0 and 167.2

1. If you choose 130, everyones baseline earning is 130 (since 130 is the minimum of the choices of all group members including yourself). As indicated in Point A, your supplementary earning will be 15. This is because your choice is 30 below 160, and 1.5 times this difference exceeds 15, leaving your positive earning at its cap of 15. As a consequence, your total earning is $130 + 15 = 145$.
2. If you choose 155, everyones baseline earning is 155. As indicated in Point B, your supplementary earning will be 7.5 (1.5 times the difference between your choice of 155 and the minimum choice of the paired members of 160). As a consequence, your total earning is $155 + 7.5 = 162.5$.
3. If you choose 160, everyones baseline earning is 160. As indicated in Point C, your supplementary earning is zero because your choice is identical to the minimum of the paired members. As a consequence, your total earning is $160 + 0 = 160$.
4. If you choose 165, everyones baseline earning is 160 (since 160 is the min-

imum of choices of all group members including yourself). As indicated in Point D, your supplementary earning will be -7.5. This is because your choice is 5 above 160, and 1.5 times this difference is 7.5. As a consequence, your total earning is $160 - 7.5 = 152.5$.

5. If you choose 190, everyones baseline earning is 160. As indicated in Point E, your supplementary earning will be -15, because 1.5 times the difference between your choice and the minimum of the paired members choices is 45 which exceeds 15. As a consequence, your total earning is $160 - 15 = 145$.

Table 1 summarizes the baseline, supplementary and total point earnings for the 5 possible scenarios (when the paired members choose 160.0 and 167.2).

Your Choice	Point Indicated in Figure 2	Your Paired Members' Minimum Choice	Baseline Earning	Supplementary Point Earning	Total Point Earning
130.0	A	160.0	130.0	+15.0	145.0
155.0	B	160.0	155.0	+7.5	162.5
160.0	C	160.0	160.0	0	160.0
165.0	D	160.0	160.0	-7.5	152.5
190.0	E	160.0	160.0	-15.0	145.0

Table 1: Summary of Earnings when the Paired Members Chose 160.0 and 167.2

EXAMPLE 2: Assume your paired members chose 120.0 and 147.8, respectively. Note that the minimum of your paired members choices is 120.0 (the smaller of 120.0 and 147.8). We shall use Figure 3 below (which we created by relabeling Figure 1) to determine your supplementary earning depending on possible choices you can make. Note that the origin is now set to l20.

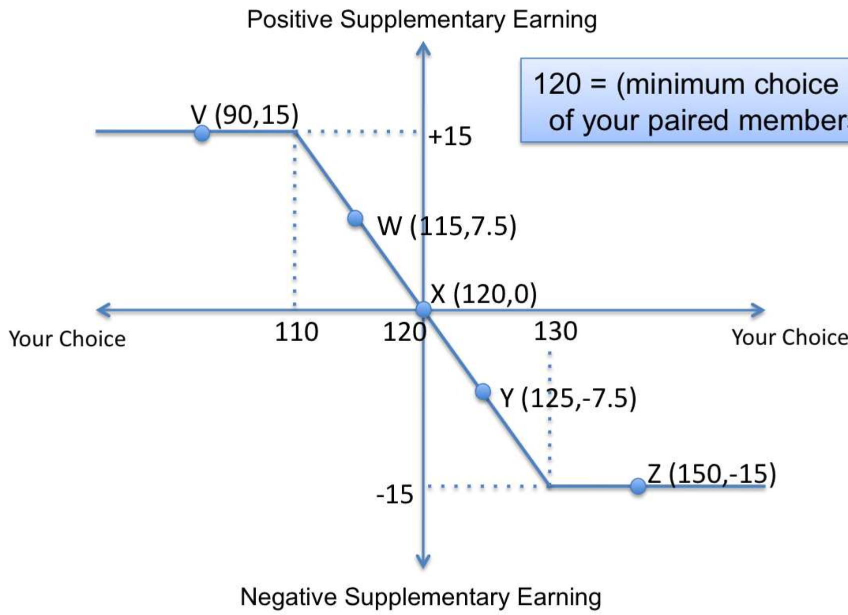

Figure 3: Choices Scenarios when the Paired Members Chose 120.0 and 147.8

1. If you choose 90, everyones baseline earning is 90 (since 90 is the minimum of the choices of all group members including yourself). As indicated in Point V, your supplementary earning will be 15. This is because your choice is 30 below 120, and 1.5 times this difference exceeds 15, leaving your positive earning at its cap of 15. As a consequence, your total earning is $90 + 15 = 105$.
2. If you choose 115, everyones baseline earning is 115. As indicated in Point W, your supplementary earning will be 7.5 (1.5 times the difference between your choice of 115 and the minimum choice of the paired members of 120). As a consequence, your total earning is $115 + 7.5 = 122.5$.
3. If you choose 120, everyones baseline earning is 120. As indicated in Point X, your supplementary earning is zero because your choice is identical to the minimum of the paired members choices. As a consequence, your total earning is $120 + 0 = 120$.

4. If you choose 125, everyones baseline earning is 120 (since 120 is the minimum of choices of all group members including yourself). As indicated in Point Y, your supplementary earning will be -7.5. This is because your choice is 5 above 120, and 1.5 times this difference is 7.5. As a consequence, your total earning is $120 - 7.5 = 112.5$.
5. If you choose 150, everyones baseline earning is 120. As indicated in Point Z, your supplementary earning will be -15 because 1.5 times the difference between your choice and the minimum of the paired members choices is 45, which exceeds 15. As a consequence, your total earning is $120 - 15 = 105$.

Table 2 summarizes the baseline, supplementary and total point earnings for the 5 possible scenarios (when the paired members choose 120 and 147.8).

Your Choice	Point Indicated in Figure 3	Your Paired Members' Minimum Choice	Baseline Earning	Supplementary Point Earning	Total Point Earning
90.0	V	120.0	90.0	+15.0	105.0
115.0	W	120.0	115.0	+7.5	122.5
120.0	X	120.0	120.0	0	120.0
125.0	Y	120.0	120.0	-7.5	112.5
150.0	Z	120.0	120.0	-15.0	105.0

Table 2: Summary of earnings when the paired members chose 120.0 and 147.8

Your Dollar Payoffs

At the end of the experiment, we will sum your point earning in each round to obtain your total point earning over 10 rounds. We will then multiply your total point earning by \$0.01 to obtain your dollar earning. In addition, we will add to this dollar earning \$5 show-up fee to determine your final dollar earning. The amount will be paid to you in cash before you leave the experiment today.