A Bayesian Cognitive Hierarchy Model with Fixed Reasoning Levels

September 2021

Jun Feng[a,b], Xiangdong Qin[c], and Xiaoyuan Wang[d,e*]
[a] School of Management, Zhejiang University of Technology, Hangzhou, China
[b] Institute of Neuromanagement, Zhejiang University of Technology, Hangzhou, China
[c] Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, China
[d] School of Management and Economics, University of Electronic Science and Technology of China, Chengdu, China
[e] Center for Israeli Studies, University of Electronic Science and Technology of China, Chengdu, China

* Corresponding author
E-mail addresses: cb8226@zjut.edu.cn (Jun Feng), xdqin@sjtu.edu.cn (Xiangdong Qin), xywang1987@gmail.com (Xiaoyuan Wang)

Abstract: We propose a Bayesian cognitive hierarchy (BCH) model with fixed reasoning levels for two-person normal-form games. The model extends the previous static version of the cognitive hierarchy model to dynamic environments and combines the cognitive hierarchy model with one of the most advanced adaptive learning models. We estimate the proposed model and other models with five datasets of two-person repeated normal-form games. The results indicate that the fixed-level BCH model can reasonably capture changes in the sophistication of behavior over time. Compared with the adaptive learning model, introducing reasoning can significantly improve the interpretation of data. We further decompose the BCH model to investigate the effect of each modeling component and find that, in different games, players rely on different decision-making processes of learning and reasoning.

Keywords: cognitive hierarchy; experience-weighted attraction; learning; level-*k* reasoning

JEL classification: C72; C91

Acknowledgment: We appreciate the helpful comments of Xing Zhang, Juin Kuan Chong, Ed Hopkins, and participants at the seventh Xiamen University International Workshop on Experimental Economics and the 2018 SJTU Experimental Economics Workshop. Xiaoyuan Wang received support from the Natural Science Foundation of China [71502023].

1 Introduction

Both the management and economics literatures document that players' choices in many games deviate systematically from equilibrium predictions (Ho et al., 2021). Level-*k* (LK, e.g., Stahl and Wilson, 1994, 1995; Nagel, 1995; Costa-Gomes et al., 2001, 2009; Crawford, 2003; Costa-Gomes and Crawford, 2006; Crawford and Iriberri, 2007) and cognitive hierarchy (CH, e.g., Camerer et al., 2004; Chong et al., 2016) models have proven to be powerful in explaining certain non-equilibrium behavior in one-shot games. In these models, players of higher reasoning levels are assumed to best respond to choices of the perceived lower-level players.[1] The estimated reasoning levels intuitively reflect differences in reasoning ability among players. However, when a game repeats over time, players not only employ limited depths of reasoning but learn from previous trials, as well. Players' learning behavior may confound the estimation of hierarchical reasoning. Therefore, it is crucial to incorporate learning into these reasoning models.

In one of the earliest explorations of the limited depth of reasoning, Nagel (1995) already reports that players learn from other players' past choices while applying level-*k* reasoning. More recently, Gill and Prowse (2016) argue that players not only learn to better respond to opponents in an adaptive way, but also learn to deepen the reasoning rules. The "rule learner" represents a more sophisticated way of learning. In Gill and Prowse (2016)'s estimation framework, rule learners move up one reasoning level over the course of the game. Ho et al. (2021) explicitly model the rule learners' beliefs in a Bayesian level-*k* model (BLK). Rule learners are assumed to update their beliefs about the distribution of opponent reasoning levels. Then, they sample one level from the belief distribution of opponent reasoning levels and noisily respond to this level. This assumption allows rule learners to choose from among all sophisticated reasoning levels (i.e., level ≥ 1). In other words, sophisticated players are homogeneous, and their reasoning levels are unbounded.

However, in real-world scenarios, such as stock investments, oligopoly competitions, and auctions, players' sophistication is likely to be bounded by their (heterogeneous) costs and the benefits of reasoning (Alaoui and Penta, 2016). Given a specific game and specific opponents, different players may "optimize" their reasoning effort and show different levels of sophistication. These sophistication levels may also vary over time, when a player updates her belief about opponents' sophistication levels. Therefore, the ability to capture heterogeneous reasoning levels and beliefs is crucial, for a model to predict repeated interactions in the real world.

[1] More precisely, the level-*k* model assumes that a level-*k* player considers all other players as level *k*-1, and the cognitive hierarchy model assumes that a level-*k* player has a belief about the distribution of all lower-level players (i.e., from level 0 to level *k*-1).

In this paper, we explore a way to model heterogeneous rule learning. Instead of allowing players to deepen their reasoning levels, we posit that players' true reasoning levels are fixed over time. Although sophisticated learners are constrained by their true learning levels, they still have the ability to update the distribution of opponent reasoning levels lower than themselves. The model is built directly upon the static Cognitive Hierarchy (CH) models (Camerer et al., 2004; Chong et al., 2016) and thus, a Bayesian Cognitive Hierarchy (BCH) model. While both the BCH and BLK models assume that players update their opponents' reasoning rule distributions, there are distinctions between the two models. First, unlike the BLK model, players' reasoning levels represent their maximum reasoning ability in the BCH model. Therefore, we introduce heterogeneity among sophisticated players. This assumption is supported by the model and tests developed by Alaoui and Penta (2016); they demonstrate heterogeneity in reasoning ability, using a cost-benefit analysis. In the BCH model, when sophisticated players with different reasoning levels observe the same outcome from opponents, they interpret the observation and update their beliefs differently. If a player realizes that other players might be more sophisticated than she had imagined before, then her maximum reasoning ability would be higher than the level that her (incorrect) beliefs implied. The purpose of the BCH model is to estimate the distribution of maximum reasoning ability among players during a session.

Second, following Chong et al. (2016), our model assumes that players make responses to the entire belief distribution of opponent reasoning levels. If a level-2 player believes that her opponent is a level-0 or level-1 player with a 50% probability, in our model, she will calculate the expected payoff of each choice based on this belief. Ho et al. (2021) use a simplifying assumption that the player will sample a level from this belief and calculate the expected payoff of each choice based on the sampled level. Therefore, in the BLK model, responses by sophisticated learners belong to the rule hierarchy of level-k reasoning, while the BCH model may capture heterogeneous beliefs (about the distribution of opponent reasoning levels) and responses of the same reasoning level. In summary, the BCH model allows sophisticated players to be heterogeneous in reasoning levels as well as in their beliefs about the distribution of opponent reasoning levels. It can be used as a test bed for understanding the interaction between reasoning levels and different types of learning.

Our model consists of three main components: adaptive learning, hierarchical reasoning, and belief updating (sophisticated learning). Based on these modeling components, (high-level) players build up a sequence of reasoning levels and construct the expected payoffs and choice probabilities for each lower level at each period. We highlight these crucial modeling components as follows:

i. The adaptive learning of level-0 players. Similar to Ho et al. (2021), level-0 players are assumed to be adaptive learners. There are many ways to model their learning behavior (Ho et al., 2021). In this paper, we assume that level-0 players follow a self-tuning experience-weighted attraction (EWA) learning process (Ho et al., 2007), which is one of the most parsimonious yet powerful models of adaptive behavior.

ii. Hierarchical reasoning and its interaction with learning. High-level players can anticipate the adaptive learning of level-0 players from every player's perspective. For example, consider a two-person normal-form game. If subject i is a level-0 player, she follows an adaptive learning process. By contrast, if subject i is a level-1 player, she needs to simulate the adaptive learning process from her opponent's viewpoint and, based on those simulated results, construct her own expected payoffs (or attractions) and choice probabilities. A level-2 player can further simulate the adaptive learning process from the viewpoints of both players and then construct the choice probabilities of her level-0 and level-1 opponents.

iii. Belief updating of high-level players. The model assumes that the relative reasoning levels of players are fixed, consistent with the original cognitive hierarchy model, but their beliefs about the relative proportions of lower-level players could vary over time. Therefore, we relax the strict assumption that the relative proportions in the minds of high-level players are the truncated Poisson distributions of the true level distribution in the previous CH model. When high-level players believe that the reasoning level of most players is relatively low, they will behave like low-level players themselves. For example, a level-3 player who knows that there are level-0, level-1, and level-2 players, and who believes most of them are level 0, will behave like a level-1 player. Conversely, if high-level players believe that the reasoning level of most players is relatively high, their own behavior will become more sophisticated as well. In other words, players are not only overconfident about their reasoning levels compared with others but can pretend to be a lower-level player to respond to an (overconfident) belief as well. [2]

Our model is designed to capture reasoning and learning in two-person repeated normal-form games with a partner-matching protocol, because high-order sophisticated behavior may be more clearly examined with fewer partners—thus fewer dimensions of anticipated learning.[3] The

---

[2] Following the original CH model (i.e., Camerer et al. (2004)), in our models, players are assumed to be strictly overconfident. i.e., level-k players treat other players as a mixture of lower levels. Unlike our approach, some studies challenge this assumption and allow players to respond to others of the same level as themselves. For example, see Koriyama and Ozkes (2021) for a discussion of the strong and weak forms of overconfidence. See also Alaoui and Penta (2016) and Levin and Zhang (2020) for alternative CH models with "self-awareness."

[3] Similar to the static CH model of Chong et al. (2016), our model can also be modified for games with more than two players. As discussed in Ho et al. (1998) and Ho et al. (2021), in n-person games, players may employ some techniques to simplify the sophisticated decision-making process. For

repeated setting enables players to learn and update beliefs about their opponents' reasoning levels, and the partner-matching setting allows players to anticipate the learning process from the viewpoint of their opponents, because players know the historical choices and observations of their opponents. We search for relevant datasets with two-person partner matching in the experimental literature and find five datasets suitable for the model, based on the following three experimental studies: Hyndman et al. (2012), O'Neill (1987), and Mookherjee and Sopher (1997) (for details, see Section 5.1).

We first compare our BCH model with the self-tuning EWA (SEWA) model and a modified BLK (hereafter, BLK*) model. The BLK* model is made by substituting the model component of cognitive hierarchy in the BCH model with the alternative assumptions of Ho et al. (2021). We find that both the BLK* and the BCH models show highly consistent estimates for the proportion of reasoning levels, and the differences in performance between the two models are data specific. Both the BCH and the BLK* models have significantly better performance compared with the SEWA model, in terms of explaining and predicting behavioral patterns over time. We further decompose the BCH model to evaluate the separate effects of adaptive learning, sophisticated learning, and hierarchical reasoning in different games. We find that the adaptive learning of level-0 players makes significant improvements in explaining and predicting behavior in three of the five datasets. Hierarchical reasoning significantly improves the interpretation and prediction in all five datasets. Bayesian updating makes a significant contribution in two of the five datasets.

The main contribution of the current research is that we elaborate on a way to integrate an established model of learning (Ho et al., 2007) and an established model of reasoning heterogeneity (Camerer et al., 2004; Chong et al., 2016). A key feature of our model is that subjects' maximum reasoning abilities are heterogeneous. We show that introducing heterogeneous reasoning abilities into the most advanced adaptive learning model can significantly improve the interpretation and prediction of data in two-person repeated normal-form games. Moreover, the BCH model achieves better or similar performance, compared with the BLK* model, indicating that even though the reasoning levels are fixed over time, the change in beliefs is sufficient to interpret the change in sophistication. With this unified framework of learning and reasoning, we show that, in different two-person normal-form games, players rely on different types of learning and reasoning: adaptive learning dominates hierarchical reasoning in games with a pure strategy Nash equilibrium, while the opposite is true in constant-sum games.

---

example, players may treat opponents as a single aggregate player. Such an extension is not considered in this paper.

2 Related Literature

Instead of completely reasoning toward an equilibrium, players usually learn adaptively in repeated games. There are two major ways to model such adaptive behavior. The first focuses on modeling players' beliefs about other players' strategies (e.g., Cheung and Friedman, 1997). Because in many repeated games, players can keep track of their opponents' past strategies, it is natural to assume that they form beliefs based on their past observations. The second way of modeling involves choice reinforcement (e.g., Roth and Erev, 1995). Players are more likely to play a previously chosen strategy that improves their payoffs. Camerer and Ho (1999) consolidate these two types of learning into an experience-weighted attraction (EWA) learning model and show that EWA learning outperforms previous single-type learning models in repeated normal-form games. Ho et al. (2007) further propose a similarly effective and yet much more parsimonious version of the EWA learning model by replacing the key parameters in the original EWA model with functions that change over time in response to experience. We incorporate this version of the EWA model into the BCH model, as well as into the BLK* model.

The learning effect in reasoning has been discussed since the earliest tests on level-k reasoning. In the first formal experimental test of p-beauty contest games, Nagel (1995) already considers anticipatory learning of other players' thinking levels. Instead of a change in reasoning levels over time, the experimental data is more consistent with players learning towards an ex post determined "optimal" target. Learning seems to happen uniformly and be independent of the reasoning process. Nagel concludes that after accounting for this form of learning, players' reasoning levels do not change significantly over time.

More recently, we see empirical evidence of different types of learning behavior across reasoning levels, indicating an interactive effect of learning and reasoning. For example, Shachat and Swarthout (2012) show that subjects anticipate learning by other players. Gill and Prowse (2016) report that more cognitively able subjects respond to the cognitive ability of their opponents, while less cognitively able ones do not. These findings suggest that a model that captures learning, reasoning, and the interplay between them is necessary to understand the evolution of bounded rationality in a dynamic environment. Ho et al. (2021) have made the first step toward this goal. They propose a Bayesian level-k model incorporating two different types of learning behavior. The lowest-level players are adaptive learners who will change their choices adaptively. Meanwhile, higher-level players are sophisticated learners who sample a reasoning level given their beliefs and respond to the sampled rule, so that they are one level above their opponents. Therefore, all high-level players have a belief about the distribution of all reasoning levels and make a noisy response to a particular reasoning level drawn from the distribution.

There are other related studies that focus on analyzing level-k reasoning over time. Danz et al. (2012) incorporate the other-regarding preference into the level-k reasoning analysis in a two-person repeated normal-form game. Their analysis does not consider that even players with the lowest reasoning levels can learn to respond better to other players, over time. Stahl (1996) combines level-k reasoning with a reinforcement learning model for the p-beauty contest game. In such a model, players recognize the reasoning rules and the corresponding choice of each rule, and they can reinforce the choice of the better-performance rule. Ho and Su (2013) further combine level-k reasoning with a belief-based learning model for a specific set of sequential games. In their model, reasoning levels become a choice set, based on which players can design their beliefs and responses, and the learning process is built upon the reasoning levels. Because the model specification requires that each reasoning level corresponds to a particular strategy profile, the model cannot be applied to most repeated normal-form games (Ho et al., 2021).[4]

3 The Fixed-Level BCH Model

Consider a finitely repeated two-person normal-form game. For each player, there is a strategy set S_i containing m_i pure strategies. That is, $S_i = \{s_i^1, s_i^2, \dots, s_i^{m_i}\}$. $s_i(t)$ denotes the observed choice of player i at period t. For player i at period t, there is a payoff function, $\pi_i(s_i(t), s_{-i}(t))$. We describe the BCH model in the following subsections.

3.1 Prior Assumptions

The players are assumed to be heterogeneous in their reasoning levels. These reasoning levels represent the heterogeneity of the maximum reasoning ability among players. Thus, their reasoning levels are fixed over time. The distribution of the reasoning levels is captured by a Poisson distribution with parameter λ (Camerer et al., 2004). The true proportion of level-k players is $f_\lambda(k) = \frac{\lambda^k}{k!} e^{-\lambda}$ and $\sum_{k=0}^{\infty} f_\lambda(k) = 1$.

Following Ho et al. (2021), the lowest level players, or level-0 players, are assumed to be adaptive learners.[5] At the beginning of experiments, a level-0 player makes random choices with equal probability. Therefore, the prior attractions of all feasible choices are the same for the level-0 player. For player i, a level-0 player, the prior attraction of choice j, denoted as $A_i^j(k = 0, t = 0)$,

[4] Even though we can identify the reasoning level of each choice in some static normal-form game and apply the model of Ho and Su (2013) to the repeated case, their model will be similar to previous belief-based learning models (e.g., the fictitious play proposed by Brown (1951) and the γ-weighted fictitious play proposed by Cheung and Friedman (1997)).

[5] Note that belief-based learning requires players to adaptively predict the opponent's choice, which may not be consistent with the definition of level 0. However, to coincide with Ho et al. (2021), and to nest the traditional settings of static level-k and CH models as special cases of our model, the lowest reasoning level in our model is defined as level 0.

is assumed to be the average of all possible outcomes.[6] That is, $A_i^j(k=0, t=0) = \frac{1}{m_i m_{-i}} \sum_{j=1}^{m_i} \sum_{j'=1}^{m_{-i}} \pi_i(s_i^j, s_{-i}^{j'})$, where the first zero corresponds to player i's reasoning level, and the second zero represents the initial stage before the first round of the game. For level-0 players at later periods, the attraction of choice j is a weighted average between the attraction of the previous period and the forgone or actual payoff at the current period, as defined in Camerer and Ho (1999). However, for players with reasoning levels greater than zero, $A_i^j(k \geq 1, t)$ should not be interpreted as attractions that were exclusively used for adaptive learners in the SEWA model. Sophisticated players are able to calculate the expected payoffs according to low-level players' attractions. In the next subsections, we slightly abuse the notation and denote the expected payoff of choice j at period t for high-level players as $A_i^j(k, t)$, where $k \geq 1$ and $t \geq 1$.[7]

When $k \geq 1$, players' prior beliefs about the distribution of opponent reasoning levels need to be specified. For $k = 1$, the player simply treats the other player's reasoning level as zero. When $k \geq 2$, the player needs to form beliefs about the proportion of level h ($h \leq k-1$). The perceived proportion of level-h ($h \leq k-1$) players in the mind of player i at period 1 is assumed to be $g_i^h(k, 1) = \frac{f_{\lambda'}(h)}{\sum_{h'=0}^{k-1} f_{\lambda'}(h')}$, where $f_{\lambda'}(h)$ follows another Poisson distribution, with parameter λ'. Therefore, this assumption shares the advantages of the (truncated) Poisson distribution as discussed in Camerer et al. (2004). Meanwhile, a different parameter gives the model more freedom to capture the bias when players initially guess true proportions. If λ' is very close to zero, high-level ($k \geq 2$) players believe that others are level-0 players. In such a case, the Bayesian updating is invalid, because only level 0 has a positive weight, and thus, the prior belief should be invariant over time. By contrast, when λ' becomes very large, a level-k player starts with an initial belief that others are more likely to be of level *k-1*. For example, for a level-3 player, when $\lambda' = 1$, her initial perceived proportions of level-0, -1, and -2 players are 40%, 40%, and 20%, respectively; when $\lambda' = 5$, the proportions become 5%, 27%, and 68%; when $\lambda' = 50$, they become 0%, 4%, and 96%. When $\lambda' \to \infty$, the initial belief of high-level players will become close to the assumption of level-k reasoning.

[6] Chong et al. (2016) propose an alternative assumption, in which those strictly dominated choices are chosen relatively rarely. The level-0 choices with non-neutral framing are also considered in Crawford and Iriberri (2007). To nest the traditional settings of static level-*k* and CH models as special cases of our model, we choose to use the neutral setting for level-0 choices.

[7] In the BCH model, the initial attractions, $A_i^j(k=0, t=0)$, for the level-0 SEWA learners must be specified; in contrast, higher-level players do not need to specify a prior value at $t = 0$; they directly update from $A_i^j(k \geq 1, t=1)$. See also the discussion in Section 3.3.

3.2 Updating Rule of Level-0 Players

We assume that level-0 players update their beliefs and make noisy responses with the adaptive procedure of self-tuning experience-weighted attraction (SEWA) (Ho et al., 2007). First, the EWA model is much generalized and combines the features of both the belief-based learning model and the reinforcement learning model (see also Camerer and Ho, 1999). Therefore, this model reflects some common traits of adaptive behavior that can be anticipated by high-level players in repeated games.[8] Second, as suggested in Ho et al. (2007), SEWA is useful as a building block for a model that includes sophistication, because it is more parsimonious than the original EWA model. Therefore, in our models, SEWA is chosen as the basis for the learning process of level-0 players.

In the following part of this subsection, we briefly review the SEWA model of Ho et al. (2007), as well as the original EWA model (Camerer and Ho, 1999). Suppose that player i is a level-0 player. The EWA models contain two core variables, which are updated over time. The first one is $N(t)$, which is interpreted as the number of observation-equivalents of past experience in Camerer and Ho (1999). Following Ho et al. (2007), the prior value, $N(0)$, is set to 1, and $N(t) = N(t-1)\varphi + 1$, where φ is a depreciation rate that measures the fractional impact of previous experience, compared with the current period.

The second one is denoted as $A_i^j(k=0,t)$, which is the attraction of choice j for level-0 player i at period t. Its prior value is defined in the previous subsection as $A_i^j(k=0,t=0)$. $A_i^j(k=0,t\geq 1)$ is updated according to the weighted sum of a depreciated experience-weighted previous attraction, $A_i^j(k=0,t-1)$, and the actual (or forgone) payoffs induced by choice j of period t, $\pi_i(s_i^j, s_{-i}(t))$.

Following the original EWA model, $A_i^j(0,t)$ can be written as:

$$A_i^j(0,t) = \frac{\varphi N(t-1) A_i^j(0,t-1) + [\delta + (1-\delta) I(s_i^j, s_i(t))]\pi_i(s_i^j, s_{-i}(t))}{N(t)}, \tag{1}$$

where $I(s_i^j, s_i(t))$ is an indicator function that equals one if $s_i^j = s_i(t)$; φ controls for the growth of attractions; and δ measures the relative weight given to forgone payoffs, compared to actual payoffs. When $\delta = 0$, only actual payoffs are reinforced, so the learning rule is like a reinforcement learning model. When $\delta = 1$, all forgone payoffs are reinforced; Camerer and Ho (1999) show that this is equivalent to a weighted fictitious play model.

---

[8] For a related discussion, please see Shachat and Swarthout (2012). The study reports that human subjects can anticipate the learning algorithm (including the EWA learning) of robots and make responses accordingly in an experiment of repeated normal-form games.

The original EWA model is made self-tuning by Ho et al. (2007) and becomes even more parsimonious. In the SEWA model, φ and δ become dynamic functions of player i's experience, $\varphi_{\mathrm{i}}(t)$ and $\delta_{ij}(t)$. First, $\varphi_{\mathrm{i}}(t)$ captures the idea that players attempt to detect changes in the learning environment. Following Ho et al. (2007), we define $H_i^{j'}(t) = \frac{\sum_{v=1}^{t} I(s_{-i}^{j'}, s_{-i}(v))}{t}$ as the historical frequency (including the last period t) of observing choice j' from player $-i$, and $R_i^{j'}(t) = I(s_{-i}^{j'}, s_{-i}(t))$ as an indicator function for observing choice j' from player $-i$ at period t. Then $\varphi_{\mathrm{i}}(t)$ can be expressed as the following:

$$\varphi_i(t) = 1 - \frac{1}{2}\sum_{j'=1}^{m_{-i}}[H_i^{j'}(t) - R_i^{j'}(t)]^2. \qquad (2)$$

In other words, players take into account the difference between the last observation and the reference (i.e., the historical frequency). A higher $\varphi_i(t) \in [0,1]$ suggests that player i's last observation is more consistent with her history play, and thus, more weight is given to the previous attraction, $A_i^j(0, t-1)$.

Second, $\delta_{ij}(t)$ captures the idea that the forgone payoffs can be divided into two groups: those larger than the actual payoff and those smaller than the actual payoff. If the forgone payoffs induced by some particular choices were larger than the actual payoff in period t, these choices might be more salient in the next period. Therefore, SEWA assigns a higher weight for the forgone payoff, if it is higher than the actual payoff.

$$\delta_{ij}(t) = \begin{cases} 1 \quad if\ \pi_i\left(s_i^j, s_{-i}(t)\right) \geq \pi_i(\,s_i(t), s_{-i}(t)), \\ \quad 0 \quad otherwise. \end{cases} \qquad (3)$$

Given Equations (1), (2), and (3), the probability of choice j made by player i (level-0) at period t is given by

$$P_i^j(0,t) = \frac{exp(\gamma_0 A_i^j(0,t-1))}{\sum_{j'=1}^{m_i} exp(\gamma_0 A_i^{j'}(0,t-1))},$$

where γ_0 is the parameter to capture sensitivities to the differences in attraction between choices for level-0 players.

3.3 Updating Rule of Level-1 Players

Level-1 players make noisy responses to the decision-making probabilities perceived for level-0 players. The noisy responses of high-level players might come from two sources. First, noise might occur when players conjecture the decision probability of lower-level players (Goeree and Holt, 2004). Second, the noise might also happen with repeated trials. Level-1 players simulate the adaptive learning process of their opponents (level 0) and form the attraction for each choice in the feasible choice set. Because level-1 players treat all other players as level 0, they do not update the distribution of reasoning levels.

Suppose that player i is a level-1 player. The expected payoff of choice j at period t is given by

$$A_i^j(1,t) = \sum_{j'=1}^{m_{-i}} \pi_i \left(s_i^j, s_{-i}^{j'}\right) P_{-i}^{j'}(0,t),$$

where $P_{-i}^{j'}(0,t)$ is the anticipated choice probability of player $-i$'s adaptive learning process. Then, the probability of choice j made by player i (level-1) at period t is given by

$$P_i^j(1,t) = \frac{exp(\gamma_H A_i^j(1,t))}{\sum_{j'=1}^{m_i} exp(\gamma_H A_i^{j'}(1,t))},$$

where γ_H is the parameter to capture sensitivities to the differences in attractions for high-level ($k \geq 1$) players. Because Ho et al. (2021) have documented significantly noisier behavior for level-0 players (from various games), following their specification, we use different parameters to capture the heterogeneity between level-0 and more sophisticated players. Since sophisticated players of different reasoning levels follow similar responding rules, we use the same parameter for the sensitivity parameter of the sophisticated players. Moreover, a single parameter for sophisticated players also helps the identification of the model.

There is a key difference in the decision-making process between level-0 players and level-1 players. Level-0 players form their choice probabilities at period t based on the updated attractions at period $t-1$. In contrast, level-1 players form their choice probabilities at period t based on the simulated choice probabilities of level-0 players at period t. Therefore, simulating choice probabilities from the opponent's perspective plays a central role in the model, which enables the interaction between learning and reasoning.

3.4 Updating Rule of Level-k ($k \geq 2$) Players

When players are at level 2 or above, they will learn from two components of the feedback in the BCH model. The first component is the updated probabilities of the choices of all lower-level players from either player's perspective. This part is similar to the updating rule of level-1 players. The second component is the relative proportions of all lower-level players. We use the Bayesian updating rule to calculate the second component.

Suppose that player i is a level-k ($k \geq 2$) player. In the BCH model, the perceived proportion of level h in the mind of player i at period t is given by,

$$g_i^h(k,t) = \rho \cdot g_i^h(k,t-1) + (1-\rho) \cdot \frac{g_i^h(k,t-1)[\sum_{j'=1}^{m_{-i}} I\left(s_{-i}^{j'}, s_{-i}(t-1)\right) P_{-i}^{j'}(h,t-1)]}{\sum_{h'=0}^{k-1} g_i^{h'}(k,t-1)[\sum_{j'=1}^{m_{-i}} I\left(s_{-i}^{j'}, s_{-i}(t-1)\right) P_{-i}^{j'}(h',t-1)]}, t \geq 2, \quad (4)$$

where $g_i^h(k,t)$ indicates the relative proportion of level-h ($h < k$) in the mind of player i, a level-k player, at period t. $I\left(s_{-i}^{j'}, s_{-i}(t-1)\right)$ is an indicator function that equals 1 if $s_{-i}^{j'} = s_{-i}(t-1)$ and 0 otherwise, and $P_{-i}^{j'}(h,t-1)$ is the anticipated choice probability of player $-i$, a level-h player, at period $t-1$. ρ is a memory decay factor. When $\rho = 1$, the prior belief about the reasoning levels is invariant. Conversely, when $\rho = 0$, the belief updating is completely driven by the Bayesian rule. The Bayesian updating is based on the observed choice of the opponent at period $t-1$ and the perceived probability of each lower-level opponent (from level 0 to level k-1) making that choice at the beginning of the period.

Given (4), the expected payoff of choice j at period t is given by

$$A_i^j(k,t) = \sum_{j'=1}^{m_{-i}} \pi_i\left(s_i^j, s_{-i}^{j'}\right) [\sum_{h=0}^{k-1} g_i^h(k,t) P_{-i}^{j'}(h,t)]. \quad (5)$$

This expected payoff function is a dynamic version of the equation (1) in Chong et al. (2016). Then, given (5), the probability of choice j, which is chosen by player i (level-k, $k \geq 2$) at period t, is given by

$$P_i^j(k,t) = \frac{exp(\gamma_H A_i^j(k,t))}{\sum_{j'=1}^{m_i} exp(\gamma_H A_i^{j'}(k,t))}.$$

3.5 Log-Likelihood Function

Based on the preceding equations, we can estimate the BCH model. The proportions of true reasoning levels do not vary during the session. The model contains five parameters: λ for the distribution of true reasoning levels, λ' for the prior belief about the distribution, ρ for the weight of the previous belief, γ_0 for the decision errors of level-0 players, and γ_H for the decision errors of high-level players.

Suppose that we observe a series of choices from player i, $\{s_i(1), s_i(2), \dots, s_i(T)\}$. The likelihood of these choices is

$$L_i = \sum_{k=0}^{\infty} f(k) L_i^k = \sum_{k=0}^{\infty} f(k) \prod_{t=1}^{T} [\sum_{j=1}^{m_i} I\left(s_i^j, s_i(t)\right) P_i^j(k,t)]. \quad (6)$$

In equation (6), a series of choices from player i are considered as one overall observation. Therefore, the probability of this overall observation for each level (i.e., L_i^k) is a product of the observed choices' probabilities of that level (i.e., $\prod_{t=1}^{T}[\sum_{j=1}^{m_i} I\left(s_i^j, s_i(t)\right) P_i^j(k,t)]$). Then, the likelihood of a series of observed choices from player i is the sum of the probabilities of all levels weighted by the proportion of each level (i.e., $\sum_{k=0}^{\infty} f(k) L_i^k$). Aggregating over n subjects, the log-likelihood of the sample is

$$LL = \sum_{i=1}^{n} lnL_i.$$

With this log-likelihood function, because the true reasoning levels of players are fixed during the experiment, a dynamic behavioral change is interpreted as changes in beliefs about the relative proportions and choice probabilities of lower-level players.

4 The Modified BLK Model

In this section, we replace the modeling component of cognitive hierarchy in our BCH model with alternative assumptions from Ho et al. (2021). We refer to the modified model as the BLK* model. By estimating the BLK* model, we hope to understand and evaluate the simplifying assumptions of the BLK model in the two-person normal-form games. Moreover, we take the BLK* model as a benchmark to evaluate the relative performance of the BCH model.

First, following Ho et al. (2021), the level-0 players account for a proportion of α, and the high-level players (i.e., level-k players with $k \geq 1$) account for a proportion of $1-\alpha$.

Second, following Ho et al. (2021), high-level players have a belief about the distribution of all reasoning levels. The prior belief of this distribution follows a Poisson distribution with a parameter λ''. The perceived proportion of level-k players in the mind of player i at period 1 is assumed to be $B_i(k,1) = \frac{\lambda''^k}{k!} e^{-\lambda''}$. Note that this prior belief is identical for all players. Therefore, the model of Ho et al. (2021) does not consider the heterogeneity in reasoning ability among high-level players.

Third, after observing the feedback in the previous period, player i can update the belief about the distribution of all reasoning levels using the Bayesian rule with memory decay factor ρ. Suppose that player i is a high-level player. The perceived proportion of level k in the mind of player i at period t is given by,

$$B_i(k,t) = \rho \cdot B_i(k,t-1) + (1-\rho) \cdot \frac{B_i(k,t-1)[\sum_{j'=1}^{m-i} I\left(s_{-i}^{j'}, s_{-i}(t-1)\right) P_{-i}^{j'}(k,t-1)]}{\sum_{k'=0}^{\infty} B_i(k',t-1)[\sum_{j'=1}^{m-i} I\left(s_{-i}^{j'}, s_{-i}(t-1)\right) P_{-i}^{j'}(k',t-1)]}, t \geq 2, \quad (7)$$

where $B_i(k,t)$ indicates the probability of level k in the mind of player i at period t. $I\left(s_{-i}^{j'}, s_{-i}(t-1)\right)$ is an indicator function that equals 1 if $s_{-i}^{j'} = s_{-i}(t-1)$ and 0 otherwise, and $P_{-i}^{j'}(k,t-1)$ is the anticipated choice probability of player $-i$, a level-k player, at period $t-1$. Equation (7) is a simplified two-person version of the original n-person updating function used in Ho et al. (2021). Because there is only one opponent, player i doesn't need to count the number of each level among opponents and only needs to update the probability of each level for the single opponent. Thus, we simply use a single-parameter weighted sum of the previous belief and the Bayesian updated result. When $\rho = 1$, the prior belief about the reasoning levels is completely sticky. Conversely, when $\rho = 0$, the belief updating is completely driven by the Bayesian rule. Equation (7) is very close to equation (4), but in equation (7), the belief of high-level players is not bounded by their reasoning ability.

Fourth, following Ho et al. (2021), high-level players sample a particular level from the belief, $B_i(k,t)$, and make noisy responses to that level. For player i, if she samples a level-k opponent, the expected payoff of choice j at period t is given by

$$A_i^j(k+1,t) = \sum_{j'=1}^{m-i} \pi_i\left(s_i^j, s_{-i}^{j'}\right) P_{-i}^{j'}(k,t). \quad (8)$$

We substitute the normal-distributed decision errors in the original model with the logit decision errors. Given (8), the probability of choice j, which is chosen by player i at period t, is given by

$$P_i^j(k+1,t) = \frac{exp(\gamma_H A_i^j(k+1,t))}{\sum_{j'=1}^{m_i} exp(\gamma_H A_i^{j'}(k+1,t))},$$

where γ_H is also the parameter of high-level players to capture sensitivities to the differences in attractions between choices.

Fifth, following Ho et al. (2021), the likelihood function of player i is given by,

$$L_i = \alpha \cdot \prod_{t=1}^{T}[\sum_{j=1}^{m_i} I\left(s_i^j, s_i(t)\right) P_i^j(0,t)]$$
$$+(1-\alpha) \cdot \prod_{t=1}^{T}\{\sum_{k=0}^{\infty} B_i(k,t)\,[\sum_{j=1}^{m_i} I\left(s_i^j, s_i(t)\right) P_i^j(k+1,t)]\},$$

where α is the probability that the player is level 0. Then, aggregating over n subjects, the log-likelihood of the sample is also

$$LL = \sum_{i=1}^{n} lnL_i.$$

Therefore, the BLK* model contains five parameters: α for the proportion of level-0 players, λ'' for the prior belief about the distribution, ρ for the weight of the previous belief, γ_0 for the decision errors of level-0 players, and γ_H for the decision errors of high-level players. A dynamic behavioral change is interpreted as either changes in beliefs about the relative proportions and choice probabilities of lower-level players or changes in the reasoning levels of sophisticated players.

5 Estimation and Result Analysis

In this section, we estimate the models using publicly available experimental datasets of two-person repeated normal-form games. Although the proposed model can be applied to datasets of all repeated normal-form games, in the estimation exercises of the current study, we focus on two kinds of normal-form games. The first is the game with a pure strategy Nash equilibrium. The second is the game without a pure strategy Nash equilibrium, for example, a constant-sum game. In the earlier literature, adaptive learning models are widely applied to analyze the experimental data of both kinds of games. However, the outcomes of adaptive learning are quite different in these two games. For instance, in a game with a unique pure strategy Nash equilibrium, which is dominance solvable, even the simplest learning process, i.e., the myopic Cournot best-response dynamics, leads to the unique pure strategy Nash equilibrium (Moulin 1984). There is a substitutive relationship between learning and reasoning because the rule hierarchy of level-k reasoning also converges to the unique Nash equilibrium in such a game. By contrast, in a constant-sum game, there is not a pure strategy equilibrium that can be attained by adaptive learning. In this sense, the relationship between learning and reasoning is not obvious. Instead, Chong et al. (2016) and other studies (e.g., Crawford and Iriberri, 2007) find that a limited depth of reasoning plays an important role in explaining observed behavior in one-shot constant-sum games or the first period of repeated constant-sum games. With the new unified framework of learning and reasoning, we expect that the estimation results may provide insights into the difference in the interaction of learning and reasoning between these two kinds of games.

Table 1

Five Experimental Datasets Used in Estimations

Experiments[a]	Number of strategies	Number of subjects	Number of periods	Number of pure strategy equilibrium	Description of games
Hyndman et al. (2012)	3x3	64	20	1	DSG[b]
	3x3	64	20	1	nDSG
O'Neill (1987)	4x4	50	105	0	Zero-Sum
Mookherjee and Sopher (1997)	4x4	40	40	0	Constant-Sum
	6x6	40	40	0	Constant-Sum

[a] In all studies, we use the data only from the treatments with partner matching. The partner matching allows players to anticipate the learning process from the viewpoint of their opponents, because players know the historical choices and observations of their opponents.

[b] DSG (nDSG) indicates that the game is (not) dominance solvable.

5.1 Datasets

We search for experimental datasets involving two-person repeated games, with appropriate monetary incentives, a nontrivial equilibrium, and relatively long periods of fixed matching. We end up finding five representative datasets from three high-quality experimental studies, as summarized in Table 1.[9] First, the experiment of Hyndman et al. (2012) is devoted to investigating adaptive learning and teaching. In the original study, the authors use a weighted belief-based learning model to detect players' adaptive learning processes. We expect that both learning and reasoning are important in explaining the data, because teaching may induce a change in the sophistication of followers over time. Second, the original study of O'Neill (1987) aims to investigate the minmax rule. One feature of O'Neill's game is that the formal rounds are relatively long, i.e., 105 rounds.[10] Therefore, the data offers more variation in the choice of strategies, which helps in the identification of our models. The data has also been used to evaluate the first-round level-*k* responses by Crawford and Iriberri (2007).[11] Third, close to

[9] The payoff matrices of these games are provided in Appendix A. We also plot observed choices of row players for each game.

[10] Subjects are also reported to play 15 training rounds.

[11] Crawford and Iriberri (2007) use O'Neill's game for between-game validation tests, instead of directly estimating the data. Therefore, their results cannot be easily compared with ours. More

O'Neill (1987), Mookherjee and Sopher (1997) examine the theoretical predictions of mixed-strategy equilibrium and adaptive learning models. Ho et al. (2007) further find that the SEWA model can reasonably interpret the data of Mookherjee and Sopher (1997). We intend to examine the performance of new models with the aforementioned datasets.

While the datasets from p-beauty contest games are usually used to estimate thinking levels, we find that the existing games do not meet our needs. Most p-beauty contest experiments involve groups of (much) more than two players and stranger matching protocols. Because players need to predict opponents' learning behavior, a game with stranger matching becomes intractable. Even there were datasets with two-person repeated p-beauty contest games, the analysis could be problematic, because choosing zero becomes a dominant strategy with two players.[12] Therefore, strategic reasoning may not explain the data well, since choosing zero is the best response to any belief. Alternatively, Chen and Krajbich (2017) propose an epiphany learning model for the two-person p-beauty contest game.

There are other games suitable for estimating thinking levels. For example, in the 11-20 game (Arad and Rubinstein, 2012), the level 0 behavior is better specified and a level-k thinking process is naturally triggered. However, the existing literature (e.g., Arad and Rubinstein, 2012; Goeree et al., 2018; Alaoui and Penta, 2016) all uses a stranger matching protocol. In particular, to rule out learning, Alaoui and Penta (2016) and Goeree et al. (2018) only provide feedback on earnings at the end of the session.

5.2 Parameter Identification and Estimation Strategy

Traditional level-k models assume the deterministic decision rule for high-level players, i.e., a level-k player best responds to level-*k-1*, and thus players' reasoning levels could be deterministically identified. However, the BCH and BLK* models employ a nondeterministic choice framework to model the decision-making processes of both adaptive and sophisticated players. When players respond stochastically, it becomes natural that they could be of multiple reasoning levels with positive probabilities. For example, in the DSG game, in the traditional level-k model with the deterministic decision rule, i.e., γ_H is positive infinity, all high-level row players with $k \geq 3$ will choose the Nash play. Nevertheless, with the nondeterministic decision rule and relatively lower γ_0 and γ_H, the probability distribution of choices will be different from

importantly, Crawford and Iriberri (2007) make specific assumptions on how level-0 players respond to the salience of the choices. Given the long periods of the game, players' initial attractions are likely to play a minor role in our models.

[12] Costa-Gomes and Crawford (2006) propose a modified two–person p-beauty contest game with a dominance solvable equilibrium. However, they use a stranger matching protocol in their experiment and different parameters of the game in each trial. In addition, there is no feedback in each period to suppress learning.

level 1 to level 20. Because the BCH model assumes that the reasoning level is fixed over time, the decision rule parameters (i.e., γ_0 and γ_H), together with the parameter of the reasoning level distribution (i.e., λ), are identified by variations in the pooled choices of different subjects at different periods.

In addition to the parameters discussed earlier, the BCH model also features parameters on the distribution of initial beliefs (i.e., λ') and the updating process of beliefs (i.e., ρ). While these parameters can regulate players' initial responses, as well as the inter-temporal dependence of players' choices, they contribute to the additional nonlinearity in the model and may introduce various plausible explanations for a particular sequence of observed behavior.[13] In other words, there could be multiple local maxima in the log-likelihood function. The estimation procedure of the model is to find the global maximum for the log-likelihood function in the reasonable parameter space. If the model can be identified in a dataset, the combination of parameters should be unique for the global maximum of the log-likelihood function. Otherwise, the model is not identified in the dataset.[14] Ho et al. (2021) use the Matlab optimization toolbox to search the local maxima in the parameter space with randomly generated initial values, and then, pick the global maximum. We use the differential evolution (DE) algorithm to directly search the global maximum in the parameter space (Storn and Price, 1997).[15] Moreover, following Ho et al. (2021)'s approach, we search from combinations of initial values and find that the best local maximum coincides with the global maximum found by the DE algorithm. Compared with BLK*, BCH reports fewer local maxima in four out of five datasets. Details are discussed and reported in Appendix B.

In the estimations of these models, we set a reasonable upper bound, 20, for the highest reasoning level. We have $0 \le \alpha \le 1$, $\lambda > 0$, $\lambda' > 0$, $0 \le \rho \le 1$, $\gamma_0 > 0$, and $\gamma_H > 0$. In Mookherjee and Sopher (1997), the experiment of each game contains two sessions with payoffs of different magnitudes. Because the higher payoff is twice as high as the lower payoff, in the estimation of the pooled data, we set γ_0 and γ_H in the data of the high-payoff sessions as half of those in the low-payoff sessions.[16] We conduct a threefold cross validation for each model in each dataset.[17]

[13] There is also interdependence in identification among parameters. For example, if λ' is very close to zero, ρ cannot be identified. If λ is very close to zero, λ', ρ, and γ_H cannot be identified.

[14] Please refer to the discussion about the observational equivalence and identification in Lewbel (2019).

[15] If there are many different combinations of parameters for the global maximum, the iterated algorithm will be pulsing across those different combinations, as long as the number of iterations is sufficiently large.

[16] We also estimate the model with separated γ_0 and γ_H in each session. The results are quite similar. Estimation codes are provided in the online supplementary documents.

[17] In the threefold cross validation, we first divide the whole sample of each dataset into three parts, for example, the first one-third of groups, the second one-third of groups, and the last one-third of groups. Then, we estimate the models with two of three parts and conduct the out-of-sample

This validation test examines whether estimates of a subgroup of subjects can be carried over to other (similar) subjects. For a robustness check, we also estimate the models with the data of the first 70% of periods (in-sample calibration) in each dataset and use these estimated parameters to predict choices in the remaining 30% of periods (out-of-sample validation). The prediction exercise checks whether our models capture the reasoning and learning dynamics, implicitly assuming that the estimated parameters are static overtime (Camerer and Ho, 1999).

The standard errors of estimates are estimated by using the jackknife method (e.g., Camerer and Ho, 1999). To check the severity of the nonidentifiability of parameters, we calculate correlations among the parameters of the models based on those jackknife estimations in each dataset (e.g., Camerer and Ho, 1999). Appendix C reports that the mean absolute correlation of each parameter with other parameters in each dataset is low or modest, indicating that the parameters are well-identified.

To show the falsifiability of the models, the estimates of each model in each dataset are used to predict the observed choices in each of all five datasets. Appendix D reports the average probability of correct predictions calculated using this procedure. We find that the estimates of each model perform best in its own datasets. The estimates in the two three-by-three games can predict behavior in each other relatively well, yet for the Zero-Sum and Constant-Sum games, these estimates perform poorly. The estimates from the Constant-Sum games outperform the strategy of choosing randomly in other similar Constant-Sum or Zero-Sum games, yet they fail to predict choices in the first two datasets.

5.3 Comparing the BCH Model with Other Models

We first report the estimation results of three models: the SEWA model, the BLK* model, and the BCH model. Table 2 lists the results of the full-sample estimation and validations of each model in the five datasets. We highlight the best outcomes in the table. Three main results can be derived from the table.

First, both the Bayesian information criterion (BIC) and validations show that the BLK* and BCH models outperform the SEWA model in all datasets. Note that the SEWA model is nested in the BLK* and BCH models. As reported in Table 2, the likelihood ratio tests, based on the full-sample log-likelihood, also shows that all of the differences in the model performance between the SEWA model and the BLK* and BCH models are significant at the 1% significance level in all five datasets. Therefore, reasoning is an important component that helps explain players' choice variations over time. Second, the BIC shows that the BLK* model dominates the BCH model in

validation using the remaining one. We repeat the procedure three times for all possible combinations and report the average of the log-likelihood values from the out-of-sample validations.

the dataset of DSG, but the difference is insignificant between the two models according to the Vuong's closeness test.[18] Furthermore, the results of both the estimation and validations that BLK* and BCH models perform similarly in the datasets of nDSG and Constant-Sum 4x4. Therefore, in these three datasets, we conclude that, compared with the BLK* model, the performance of the BCH model is quite similar. Third, we find that the BCH model significantly outperforms the BLK* model in terms of the full-sample estimation in the datasets of Zero-Sum and Constant-Sum 6x6. Overall, the BCH model performs best in two of the five datasets and similarly to the BLK* model in the other three datasets.

Table 2
Model Fitting and Validation

Datasets	DSG	nDSG	Zero-Sum	Constant-Sum	
	3x3	3x3	4x4	4x4	6x6
Estimation results with the full sample					
Obs.	1280	1280	5250	1600	1600
Log-likelihood					
SEWA	-942.54	-1014.91	-7278.02	-2132.51	-2777.10
BLK*	**-871.11**	-974.79	-6979.34	**-1931.35**	-2444.64
BCH	-903.12	**-969.61**	**-6918.06**	-1932.62	**-2407.72**
Compared to SEWA (χ^2, dof, p-value)					
BLK*	(142.86, 4, 0.00)	(80.24, 4, 0.00)	(597.36, 4, 0.00)	(402.32, 4, 0.00)	(664.92, 4, 0.00)
BCH	(78.84, 4, 0.00)	(90.60, 4, 0.00)	(719.92, 4, 0.00)	(399.78, 4, 0.00)	(738.76, 4, 0.00)
BLK* vs. BCH (Z, n, p-value)					
Vuong test	(1.26, 64, 0.21)	(-1.32, 64, 0.19)	(-2.51, 50, 0.01)	(0.09, 40, 0.93)	(-2.49, 40, 0.01)
Bayesian information criterion (BIC)[a]					
SEWA	1892.23	2036.97	14564.61	4272.40	5561.58
BLK*	**1777.99**	1985.35	14001.51	**3899.59**	4926.17
BCH	1842.01	**1974.99**	**13878.95**	3902.13	**4852.33**
Threefold cross-validation					
Average log-likelihood					
SEWA	-331.00	-354.81	-2330.88	-747.51	-975.15
BLK*	**-310.51**	-340.09	-2241.42	**-676.19**	-865.50
BCH	-318.88	**-339.13**	**-2217.86**	-676.98	**-852.41**
Out-of-sample validation with the last 30% of periods					
Log-likelihood					
SEWA	-229.77	-246.73	-2148.76	-633.42	-837.57
BLK*	**-222.02**	**-236.14**	-2056.98	-596.70	**-739.05**

[18] In the Vuong's closeness tests of this subsection, we take a series of observations from a subject as one independent observation.

BCH	-228.98	-237.14	**-2050.84**	**-580.21**	-743.41

[a] BIC (Bayesian Information Criterion) is given by $klog(N) - 2LL$, where k is the number of parameters, and N is the number of observations.

The threefold cross-validation shows that the relative performance of BCH and BLK* models, in terms of the average log-likelihood values, is consistent with their estimation performance. The out-of-sample validation with the last 30% periods shows that the BCH model achieves higher log-likelihood values in the Zero-Sum and Constant-Sum 4x4. Yet in the Constant-Sum 6x6 game, the BLK* model performs better in the hold-out sample.[19]

Table 3 shows the estimates of parameters of the models. First, the variation of estimates of $\hat{\alpha}$ in the BLK* model among the five datasets is consistent with that of the estimates of $\hat{\lambda}$ in the BCH model, but in all five datasets, the BLK* model suggests more level-0 players than the BCH model does. In the first two datasets, the BLK* (BCH) model suggests that around 64.1% (59.6%) in DSG and 58.2% (54.6%) in nDSG players are level-0 players. In the dataset of the Zero-Sum game, 20.3% (10.4%) of players are level-0 players, by the BLK* (BCH) model. In the two Constant-Sum games, the BLK* (BCH) model indicates that level-0 players account for 15.4% (1.7%) in the Constant-Sum 4x4 and 5.3% (2.6%) in the Constant-Sum 6x6. In brief, both the BCH and BLK* models suggest that most players are sophisticated players, in the games without a pure strategy equilibrium. Conversely, in the first two games with a pure strategy equilibrium, most players are adaptive learners. In a simulation exercise, we find that the adaptive learning of level-0 players converges to the Nash equilibrium quickly (i.e., choosing the equilibrium strategy with the highest probability) in the first two games. This may explain why most players are adaptive learners, because they do not need to achieve high levels of reasoning to reach the Nash equilibrium.

Second, there are also some inconsistent interpretations between the BCH and BLK* models. For example, in the dataset of nDSG, the BLK* model suggests that all high-level players believe that other players are level-0 players (i.e., $\widehat{\lambda''} = 0$). Therefore, the dataset only contains two levels, i.e., level 0 and level 1. By contrast, although the proportion of high-level ($k \geq 2$) players is around 12%, the BCH model suggests that the belief of high-level ($k \geq 2$) players is close to the assumption of level-k reasoning (e.g., a level-2 player believes that others are level 1).[20]

[19] The stationarity assumption of parameters may not hold (Camerer and Ho, 1999). It is possible that in the last 30% of periods, a subject realizes that her opponent is more sophisticated than she had imagined before, and then, her true reasoning level should be higher than the level she appeared to demonstrate in the first 70% of periods.

[20] In estimation, the upper bound for reasoning levels is set at 20. However, to estimate the distribution of players' prior belief more flexibly, we allow λ' to reach an even higher number. When λ' is sufficiently high, it suggests that the initial belief of a high-level player (e.g. $2 \leq k \leq 5$) is that the opponent is mostly one level below him.

Furthermore, in all three games without pure strategy equilibrium, the BLK* model shows that the prior beliefs of high-level players are invariant (i.e., $\hat{\rho} = 1$), but the BCH model suggests that there exists a partially Bayesian updating process (the likelihood ratio test in the next subsection shows that $\hat{\rho}$ is significantly different from one in the two Constant-Sum games). For instance, in the Constant-Sum 4x4 game, players weight the initial priors by 0.851^t after t periods. We consider this observation a crucial difference between the two models. The BCH model captures the change of sophistication in the data by interpreting it as the change in beliefs rather than the change in levels.

Table 3
Parameter Estimates with the Full Sample

Datasets	DSG	nDSG	Zero-Sum	Constant-Sum	
	3x3	3x3	4x4	4x4	6x6
SEWA					
$\widehat{\gamma_0}$	0.045	0.045	0.002	0.279	0.272
	(0.001)	(0.001)	(0.002)	(0.008)	(0.008)
BLK*					
$\hat{\alpha}$	0.641	0.582	0.203	0.154	0.053
	(0.016)	(0.013)	(0.016)	(0.015)	(0.009)
$\widehat{\lambda'}$	2.569	0.000	1.273	1.916	1.468
	(0.143)	(0.000)	(0.195)	(0.042)	(0.028)
$\hat{\rho}$	0.813	<u>1</u>[a]	1.000	1.000	1.000
	(0.013)	-	(0.000)	(0.000)	(0.000)
$\widehat{\gamma_0}$	0.042	0.063	0.091	0.053	0.000
	(0.000)	(0.001)	(0.020)	(0.003)	(0.000)
$\widehat{\gamma_H}$	6.900	0.051	0.434	1.467	1.138
	(0.421)	(0.003)	(0.004)	(0.048)	(0.011)
BCH					
$\hat{\lambda}$	0.517	0.606	2.267	4.078	3.664
	(0.043)	(0.026)	(0.060)	(0.193)	(0.183)
$\widehat{\lambda'}$	18.647	50.000[b]	2.756	2.395	1.549
	(2.338)	(8.995)	(0.079)	(0.065)	(0.075)
$\hat{\rho}$	0.000	1.000	0.978	0.851	0.861
	(0.000)	(0.177)	(0.006)	(0.014)	(0.022)
$\widehat{\gamma_0}$	0.068	0.067	0.135	0.000	0.000
	(0.001)	(0.001)	(0.006)	(0.000)	(0.000)
$\widehat{\gamma_H}$	0.029	0.056	0.391	1.128	1.251
	(0.003)	(0.002)	(0.006)	(0.015)	(0.017)

[a] In the estimation of BLK* with the data of nDSG, the estimate of $\widehat{\lambda'}$ is very close to the lower bound, 0. Therefore, the Bayesian updating is invalid in such a case. We simply set $\hat{\rho} = 1$, because the prior belief is

invariant over time.

[b] In the estimation of BCH with the data of nDSG, the estimate of $\widehat{\lambda'}$ reaches the upper bound, 50, which indicates that, for most high-level players (e.g. $2 \leq k \leq 5$), the distribution of prior beliefs is very close to the assumption of level-k reasoning.

The estimation results of the BCH model are also intuitive and coincide with the findings in Hyndman et al. (2012). For example, Hyndman et al. (2012) find that, in initial periods, players choose the Nash choice to teach their opponents where the Nash play is. In the interpretation of the BCH model, this is equivalent to suggesting that players think others are more sophisticated than they really are in the initial periods, and thus, the estimate of $\widehat{\lambda'}$ is relatively large in the DSG and nDSG games. However, the estimation results in the two constant-sum games are different from the findings reported in Mookherjee and Sopher (1997) and Ho et al. (2007). The results of the BCH model in the two constant-sum games show that the adaptive learning is negligible and hierarchical reasoning with sophisticated learning can achieve better explanations and predictions than the pure adaptive learning models.

Overall, we have two main conclusions from the estimation results in this subsection. First, the BCH model can reasonably capture the change in the sophistication of behavior, and the difference between the BLK* and BCH models is data specific. As highlighted in Ho et al. (2021), the simplifying assumption of best responding to a random draw ensures that the best response belongs to the rule hierarchy. Therefore, if an observed choice departs from the rule hierarchy, the observed choice can only be interpreted as a decision error around the closest reasoning rule by the BLK* model. By contrast, the BCH model can find a reasonable reasoning level falling outside the rule hierarchy for the observed choice via adjusting the belief updating process. In this sense, the BCH model may have more explanatory power than the BLK* model in part of the data. Second, more importantly, compared to the most advanced pure adaptive learning model, the introduction of reasoning can obtain significantly more explanatory power to the data. On the basis of the SEWA model, after introducing reasoning in two ways, both the BLK* and BCH models report significant improvements in all five datasets.

5.4 Effects of Adaptive Learning, Hierarchical Reasoning, and Bayesian Updating

The BCH model consists of three modeling components: adaptive learning, hierarchical reasoning, and Bayesian updating. In this subsection, we intend to check the marginal contribution of each modeling component. Similar to Ho et al. (2021), we also report the estimation results of three nested models. The first is a <u>BCH without SEWA</u> model (hereafter, BWS), in which we extract the SEWA learning from the BCH model by setting $\gamma_0 = 0$. Then, level-0 players make random choices with equal probability in each period, and thus, the effect of adaptive learning is removed.[21] The second is a BCH without Bayesian updating model (hereafter, BWB), in which we remove the Bayesian updating process from the BCH model by fixing $\rho = 1$. The third is a static CH model (SCH), whereby we remove the effects of both adaptive learning and Bayesian updating by setting $\gamma_0 = 0$ and $\rho = 1$.

Table 4

Model Fitting and Validation of the Three Nested Models

Datasets	DSG	nDSG	Zero-Sum	Constant-Sum	
	3x3	3x3	4x4	4x4	6x6
Estimation results with the full sample					
Obs.	1280	1280	5250	1600	1600
Log-likelihood					
BWS	-1076.86	-1344.63	-6959.03	-1932.62	-2407.72
BWB	-903.55	-969.61	-6918.73	-1940.13	-2423.13
SCH	-1079.39	-1349.87	-6959.15	-1940.13	-2423.13
Compared with the BCH model (χ^2, dof, p-value)					
BWS	(347.48, 1, 0.00)	(750.04, 1, 0.00)	(81.94, 1, 0.00)	(0.00, 1, 1.00)	(0.00, 1, 1.00)
BWB	(0.86, 1, 0.35)	(0.00, 1, 1.00)	(1.34, 1, 0.25)	(15.02, 1, 0.00)	(30.82, 1, 0.00)
SCH	(352.54, 2, 0.00)	(760.52, 2, 0.00)	(82.18, 2, 0.00)	(15.02, 2, 0.00)	(30.82, 2, 0.00)
BWB vs. SEWA (χ^2, dof, p-value)					
	(77.98, 3, 0.00)	(90.60, 3, 0.00)	(718.58, 3, 0.00)	(384.76, 3, 0.00)	(707.94, 3, 0.00)
Bayesian Information Criterion (BIC)					
BWS	2182.34	2717.88	13952.32	3894.75	4844.95
BWB	1835.72	1967.84	13871.72	3909.77	4875.77
SCH	2180.24	2721.20	13944.00	3902.39	4868.39
Threefold cross-validation					
Average log-likelihood					
BWS	-385.46	-463.78	-2226.41	-677.01	-852.41
BWB	-318.59	-338.25	-2217.60	-678.77	-854.51
SCH	-385.22	-467.82	-2226.33	-678.77	-854.51
Out-of-sample validation with the last 30% periods					
Log-likelihood					
BWS	-336.39	-420.97	-2052.82	-580.21	-743.41
BWB	-227.88	-236.41	-2045.49	-584.91	-739.65
SCH	-342.11	-422.11	-2051.44	-584.91	-739.65

[21] Meanwhile, high-level players are still able to update their perceived level distribution according to the observed play of lower-level players. γ_H still captures the precision of the decision according to the rule.

Table 4 reports the estimation results of these three nested models. To compare with the BCH model, the results of likelihood ratio tests are also provided. We can derive four main conclusions from these results. First, via comparing the BWS model with the BCH model, in the first three datasets, introducing adaptive learning can significantly improve the performance in explaining and predicting the data, but in the latter two datasets, the contribution of adaptive learning is negligible. Second, the results of the BWB model conversely show that the contribution of Bayesian updating is minor in the first three datasets, but significant in the latter two datasets. Third, to evaluate the marginal effects of hierarchical reasoning, we can compare the BWB model with the SEWA model. It also shows that adding hierarchical reasoning to the SEWA model significantly improves explanatory and predictive power in all datasets. Fourth, the estimation results from the SCH model further show that, if we remove both adaptive learning and Bayesian updating from the BCH model, the performances become significantly worse in all five datasets.

The parameter estimates of the three nested models are provided in Table 5. We note the following three patterns. First, the estimates of $\hat{\lambda}$ are variable across models. The average of fixed levels in these models is prone to be affected by adaptive learning or Bayesian updating. Second, the estimates of $\widehat{\lambda'}$ are less variable across models in the three games without pure strategy equilibrium than the first two games ($\widehat{\lambda'}$ ranges from 1.752 to 50.000 in the first two datasets but from 1.549 to 3.220 in the latter three datasets). Third, except for the first dataset, the estimate of $\widehat{\gamma_H}$ does not vary much across models. According to the estimates of both the BLK* and BCH models, most players are of level 0 in the first two games, but only a few level-0 players exist in the three games without a pure strategy equilibrium. Removing adaptive learning has a stronger impact on these estimates in the first two games than in the latter three games.

In summary, in the datasets of Hyndman et al. (2012) and the dataset of O'Neill (1987), we find that both adaptive learning and hierarchical reasoning are important in explaining and predicting the data, but the effect of Bayesian updating is minor. However, in the datasets of Mookherjee and Sopher (1997), both hierarchical reasoning and Bayesian updating play major roles in the behavior of players, but adaptive learning becomes negligible. Furthermore, in terms of the effect size, adaptive learning dominates hierarchical reasoning in the first two datasets, but the opposite is true in the latter three datasets. The fundamental differences in those selected games may contribute to the differences in players' decision-making processes of learning and reasoning picked up by the BCH model. When Ho et al. (2021) apply the original BLK model to various games (including a beauty contest game, a Cournot competition game, and an auction game), authors argue that the "convergence patterns" of the games lead to differences in sophisticated versus adaptive learning. Koriyama and Ozkes (2021) also suggest that for games in which convergence to Nash equilibrium is not guaranteed, players rely on more sophisticated learning. In a similar vein, we find that the key difference between the selected games involves the convergence to a specific pure strategy (versus mixing from a set of strategies). When players cannot settle on a specific strategy, it becomes reasonable that they engage in a higher level of reasoning, as well as sophisticated learning. Another difference between the datasets is that the games with a unique pure strategy equilibrium have fewer strategies, making these games relatively easier to comprehend. Therefore, less sophisticated reasoning may be required for them.

Table 5
Parameter Estimates of the Three Nested Models

Datasets	DSG	nDSG	Zero-Sum	Constant-Sum	
	3x3	3x3	4x4	4x4	6x6
BWS ($\gamma_0 = 0$)					
$\hat{\lambda}$	2.576	0.969	4.247	4.078	3.664
	(0.114)	(0.039)	(0.104)	(0.193)	(0.183)
$\widehat{\lambda'}$	21.790	1.752	3.220	2.395	1.549
	(1.035)	(0.227)	(0.076)	(0.065)	(0.075)
$\hat{\rho}$	0.997	0.000	0.993	0.851	0.861
	(0.001)	(0.000)	(0.003)	(0.014)	(0.022)
$\widehat{\gamma_H}$	0.606	0.068	0.384	1.128	1.251
	(0.061)	(0.001)	(0.003)	(0.015)	(0.017)
BWB ($\rho = 1$)					
$\hat{\lambda}$	0.489	0.606	2.310	4.355	2.492
	(0.040)	(0.026)	(0.061)	(0.266)	(0.063)
$\widehat{\lambda'}$	50.000[a]	50.000[a]	2.827	2.448	3.133
	(0.006)	(0.014)	(0.063)	(0.054)	(0.094)
$\widehat{\gamma_0}$	0.067	0.067	0.139	0.000	0.000
	(0.001)	(0.001)	(0.006)	(0.000)	(0.000)
$\widehat{\gamma_H}$	0.027	0.056	0.381	1.040	1.214
	(0.003)	(0.002)	(0.003)	(0.011)	(0.013)
SCH ($\gamma_0 = 0$ and $\rho = 1$)					
$\hat{\lambda}$	2.304	0.809	4.318	4.355	2.492
	(0.104)	(0.035)	(0.097)	(0.266)	(0.063)
$\widehat{\lambda'}$	23.488	50.000[a]	3.209	2.448	3.133
	(0.801)	(0.000)	(0.071)	(0.054)	(0.094)
$\widehat{\gamma_H}$	0.654	0.064	0.380	1.040	1.214
	(0.043)	(0.001)	(0.003)	(0.011)	(0.013)

[a] The estimate of $\widehat{\lambda'}$ reaches the upper bound, 50.

6 Discussion

In this paper, we propose a Bayesian cognitive hierarchy model with fixed reasoning levels. Players first update their beliefs across periods, and then, based on the result of learning, they go through a reasoning process within each period and adjust their beliefs about the relative proportion of lower levels. Differing from the most advanced model of rule learners, our new model introduces the heterogeneity of reasoning ability among high-level players and generalizes the static cognitive hierarchy model to a dynamic version. Estimation exercises with several experimental datasets show that our new model, with fixed reasoning levels, can also reasonably capture the learning process, the reasoning process, and the interplay between them in two-person repeated normal-form games.

In the current study, we choose the SEWA model as the adaptive learning process of level-0 players. The estimation results further demonstrate that on the basis of adaptive learning, introducing reasoning by either the BLK* or BCH model can significantly improve the explanatory power in all datasets. This result can be viewed as the key evidence that reasoning plays an equally important, if not more important, role as learning does in repeated normal-form games. Because the SEWA model is one of the most advanced adaptive learning models in the literature, all the adaptive learning behaviors have been largely captured, and thus, it leaves limited space for possible misinterpretation.[22]

Our estimation results indicate that both adaptive learning and hierarchical reasoning are important in interpreting the data. However, adaptive learning alone does not explain players' choices well. Instead, hierarchical reasoning seems to be a necessary part of their decision-making processes. In some games with a pure strategy equilibrium, players may rely more on adaptive learning. Once level-0 players reach the equilibrium through adaptive learning, all high levels will choose the equilibrium responses as well. In such games, adaptive learning can greatly save the cognitive cost of reasoning. By contrast, in some other games without a pure strategy equilibrium, there is no focal point (i.e., the pure strategy equilibrium) for the adaptive players to reach, and thus, these players have to spend more cognitive resources on reasoning.

In the field of model-based functional magnetic resonance imaging (fMRI) studies, Zhu et al. (2012) and Coricelli and Nagel (2009), respectively, have checked the neural representations of adaptive learning (i.e., the EWA model) and hierarchical reasoning (i.e., the CH model). Our model assumes interactions between the two systems of adaptive learning and hierarchical reasoning. We do not claim that our results discover the mechanism of the interplay between learning and reasoning; nor do we claim that our model is better than all previous approaches.

[22] If adaptive learning is not fully captured, some adaptive learning may be misinterpreted as sophisticated learning or reasoning by models.

However, our model provides an intuitive way to unify two important nonequilibrium approaches. While we still do not have conclusive evidence of the interactions between learning and reasoning, the mechanism of players' allocations of their cognitive resources between the two systems can be further investigated using the interdisciplinary approaches of neuroeconomics. We leave this for future research.

References

Alaoui, L., & Penta, A. (2016). Endogenous depth of reasoning. *The Review of Economic Studies*, *83*(4), 1297–1333.

Arad, A., & Rubinstein, A. (2012). The 11-20 money request game: A level-k reasoning study. *American Economic Review*, *102*(7), 3561–73.

Brown, G. W. (1951). Iterative solution of games by fictitious play. *Activity Analysis of Production and Allocation*, *13*(1), 374–376.

Camerer, C., & Ho, T. H. (1999). Experience‐weighted attraction learning in normal-form games. *Econometrica*, *67*(4), 827–874.

Camerer, C. F., Ho, T. H., & Chong, J. K. (2004). A cognitive hierarchy model of games. *The Quarterly Journal of Economics*, *119*(3), 861–898.

Chen, W. J., & Krajbich, I. (2017). Computational modeling of epiphany learning. *Proceedings of the National Academy of Sciences*, *114*(18), 4637–4642.

Cheung, Y. W., & Friedman, D. (1997). Individual learning in normal form games: Some laboratory results. *Games and Economic Behavior*, *19*(1), 46–76.

Chong, J. K., Ho, T. H., & Camerer, C. (2016). A generalized cognitive hierarchy model of games. *Games and Economic Behav*ior, *99*, 257–274.

Coricelli, G., & Nagel, R. (2009). Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proceedings of the National Academy of Sciences*, *106*(23), 9163–9168.

Costa-Gomes, M. A., & Crawford, V. P. (2006). Cognition and behavior in two-person guessing games: An experimental study. *The American Economic Review*, *96*(5), 1737–1768.

Costa-Gomes, M. A., Crawford, V. P., & Iriberri, N. (2009). Comparing models of strategic thinking in Van Huyck, Battalio, and Beil's coordination games. *Journal of the European Economic Association*, *7*(2–3), 365–376.

Costa-Gomes, M., Crawford, V. P., & Broseta, B. (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica*, *69*(5), 1193–1235.

Crawford, V. P. (2003). Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions. *The American Economic Review*, *93*(1), 133–149.

Crawford, V. P., & Iriberri, N. (2007). Fatal attraction: Salience, naivete, and sophistication in experimental “hide-and-seek” games. *The American Economic Review*, *97*(5), 1731–1750.

Danz, D. N., Fehr, D., & Kübler, D. (2012). Information and beliefs in a repeated normal-form game. *Experimental Economics*, *15*(4), 622–640.

Gill, D., & Prowse, V. (2016). Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis. *Journal of Political Economy*, *124*(6), 1619–1676.

Goeree, J. K., & Holt, C. A. (2004). A model of noisy introspection. *Games and Economic Behavior*, *46*(2), 365–382.

Goeree, J. K., Louis, P., & Zhang, J. (2018). Noisy introspection in the 11-20 game. *The Economic Journal*, *128*(611), 1509–1530.

Ho, T. H., & Su, X. (2013). A dynamic level-k model in sequential games. *Management Science*, *59*(2), 452–469.

Ho, T. H., Camerer, C., & Weigelt, K. (1998). Iterated dominance and iterated best response in experimental “p-beauty contests.” *The American Economic Review*, *88*(4), 947–969.

Ho, T. H., Camerer, C. F., & Chong, J. K. (2007). Self-tuning experience weighted attraction learning in games. *Journal of Economic Theory*, *133*(1), 177–198.

Ho, T. H., Park, S. E., & Su, X. (2021). A Bayesian level-k model in n-person games. *Management Science, 67(3), 1622-1638*.

Hyndman, K., Ozbay, E. Y., Schotter, A., & Ehrblatt, W. Z. E. (2012). Convergence: An experimental study of teaching and learning in repeated games. *Journal of the European Economic Association*, *10*(3), 573–604.

Koriyama, Y., & Ozkes, A. I. (2021). Inclusive cognitive hierarchy. *Journal of Economic Behavior & Organization*, *186*, 458–480.

Lewbel, A. (2019). The identification zoo: Meanings of identification in econometrics. *Journal of Economic Literature*, *57*(4), 835–903.

Levin, D, and Zhang, L. (2020). Bridging level-k to Nash equilibrium. *The Review of Economics and Statistics*, 1-44.

Mookherjee, D., & Sopher, B. (1997). Learning and decision costs in experimental constant sum games. *Games and Economic Behavior*, *19*(1), 97–132.

Moulin, H. (1984). Dominance solvability and Cournot stability. *Mathematical social sciences*, *7*(1), 83–102.

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *The American Economic Review*, *85*(5), 1313–1326.

O'Neill, B. (1987). Nonmetric test of the minimax theory of two-person zerosum games. *Proceedings of the National Academy of Sciences*, *84*(7), 2106–2109.

Roth, Alvin E., and Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior 8*(1), 164–212.

Shachat, J., & Swarthout, J. T. (2012). Learning about learning in games through experimental control of strategic interdependence. *Journal of Economic Dynamics and Control*, *36*(3), 383–402.

Stahl, D. O. (1996). Boundedly rational rule learning in a guessing game. *Games and Economic Behavior*, *16*(2), 303–330.

Stahl, D. O., & Wilson, P. W. (1994). Experimental evidence on players' models of other players. *Journal of economic behavior & organization*, *25*(3), 309–327.

Stahl, D. O., & Wilson, P. W. (1995). On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior*, *10*(1), 218–254.

Storn, R., & Price, K. (1997). Differential evolution: A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, *11*(4), 341–359.

Zhu, L., Mathewson, K. E., & Hsu, M. (2012). Dissociable neural representations of reinforcement and belief prediction errors underlie strategic learning. *Proceedings of the National Academy of Sciences*, *109*(5), 1419–1424.

Appendix

A. The Five Games:

Game 1: Hyndman et al. (2012) – DSG

Payoff matrix:

	A1	A2	A3
A1	51, 30	35, 43	93, 21
A2	35, 21	25, 16	32, 94
A3	68, 72	45, 69	13, 62

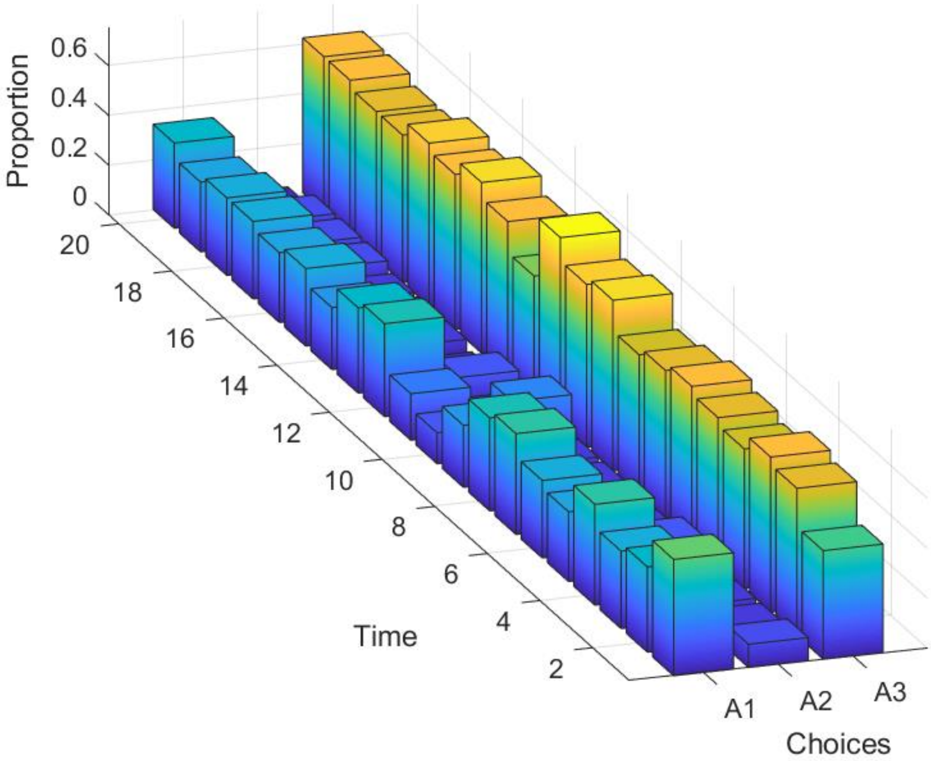

Fig. A1 Observed Choices of Row Players in DSG

Game 2: Hyndman et al. (2012) – nDSG

Payoff matrix:

	A1	A2	A3
A1	12, 83	39, 56	42, 45
A2	24, 12	12, 42	58, 76
A3	89, 47	33, 94	44, 59

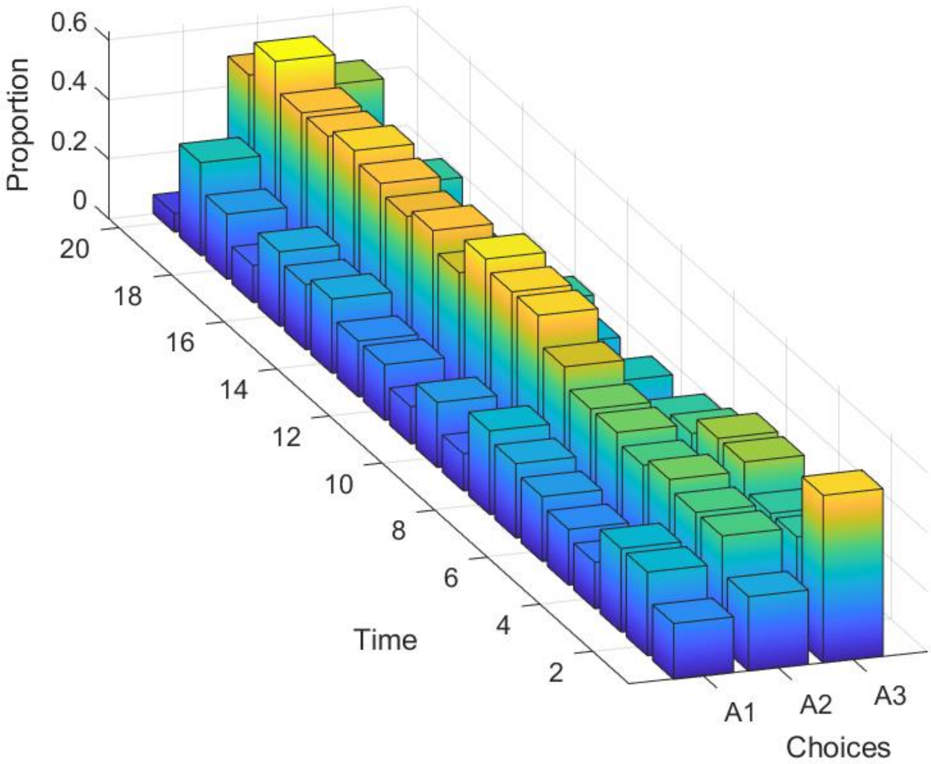

Fig. A2 Observed Choices of Row Players in nDSG

Game 3: O'Neill (1987) – Zero-Sum

Payoff matrix:

	Card Joker--1	Card Ace--2	Card 2--3	Card 3--4
Card Joker--1	5, -5	-5, 5	-5, 5	-5, 5
Card Ace--2	-5, 5	-5, 5	5, -5	5, -5
Card 2--3	-5, 5	5, -5	-5, 5	5, -5
Card 3--4	-5, 5	5, -5	5, -5	-5, 5

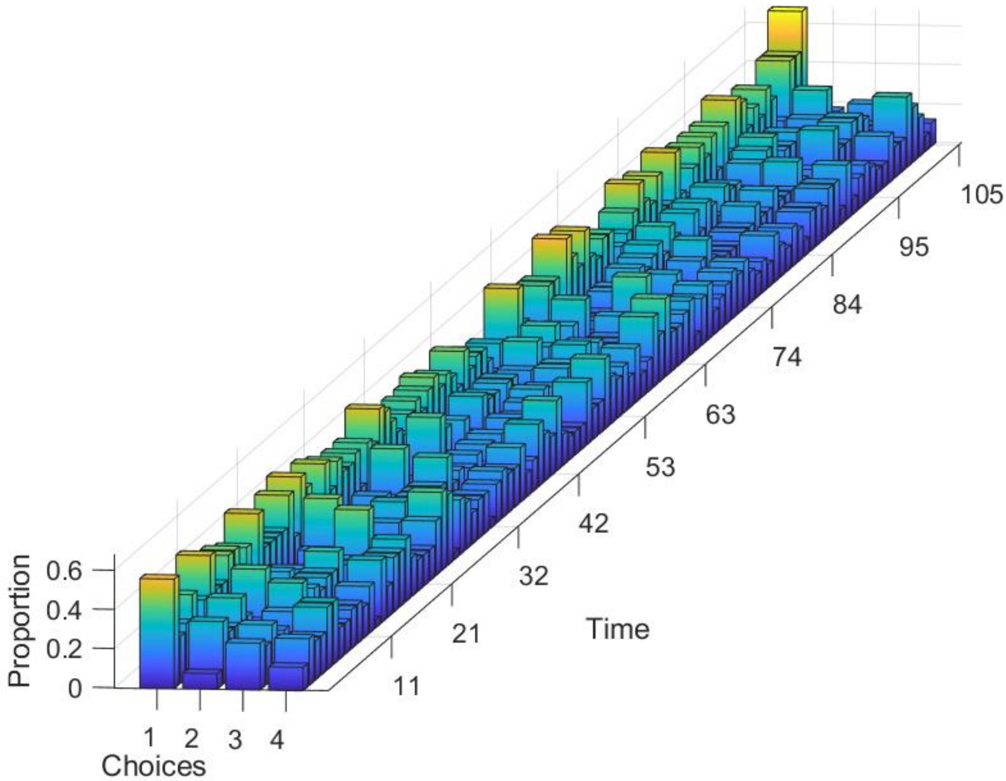

Fig. A3 Observed Choices of Row Players in Zero-Sum

Game 4: Mookherjee and Sopher (1997) – Constant-Sum 4x4

	1	2	3	4
1	x, 0	0, x	0, x	x, 0
2	0, x	0, x	x, 0	x, 0
3	0, x	x, 0	x/3, 2x/3	x/3, 2x/3
4	0, x	0, x	2x/3, x/3	x, 0

where x = 5 for session 1 and = 10 for session 3.

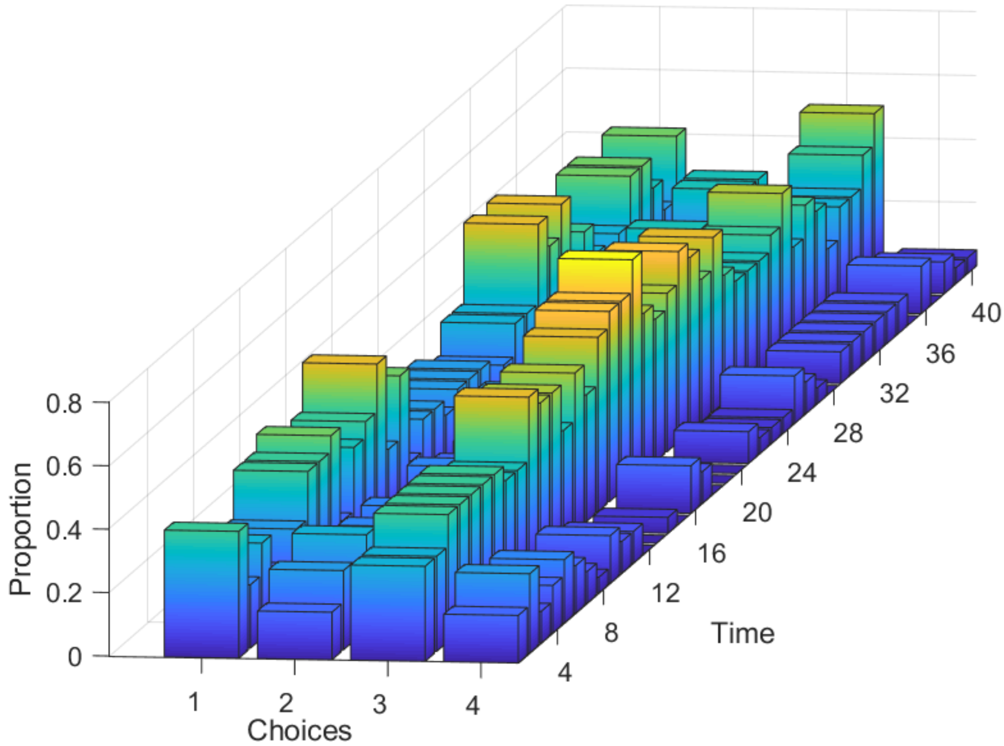

Fig. A4 Observed Choices of Row Players in Constant-Sum 4x4

Game 5: Mookherjee and Sopher (1997) – Constant-Sum 6x6

	1	2	3	4	5	6
1	x, 0	0, x	0, x	0, x	0, x	x, 0
2	0, x	0, x	x, 0	x, 0	x, 0	x, 0
3	0, x	x, 0	0, x	0, x	x, 0	0, x
4	0, x	x, 0	x, 0	0, x	0, x	0, x
5	0, x	x, 0	0, x	x, 0	0, x	x, 0
6	0, x	0, x	x, 0	0, x	x, 0	x, 0

where x = 5 for session 2 and = 10 for session 4.

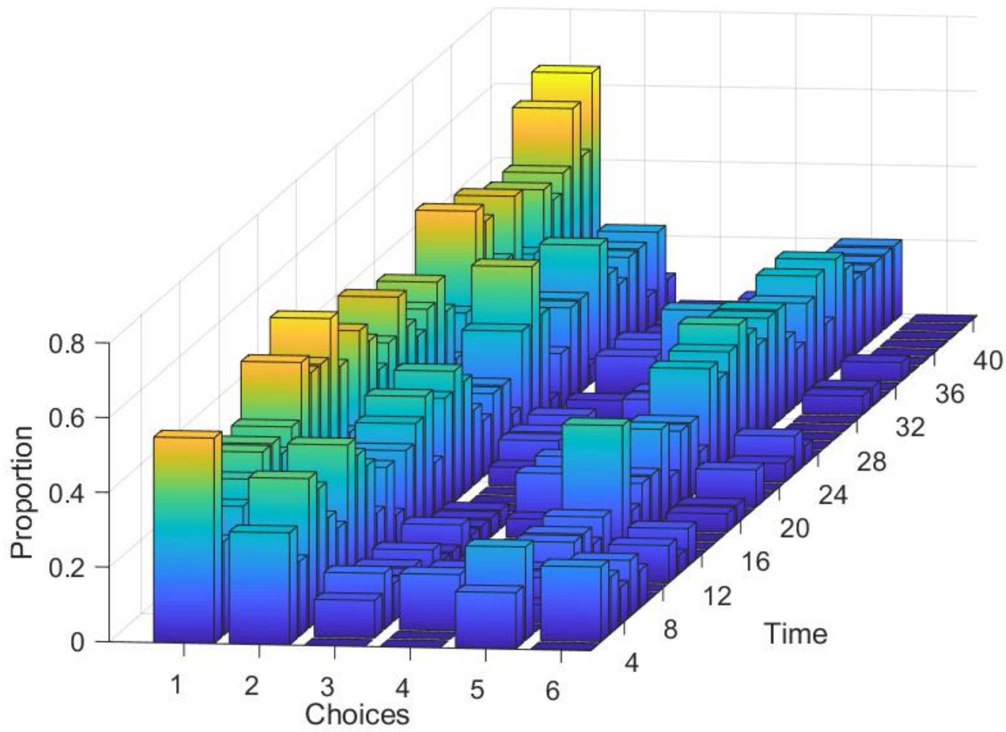

Fig. A5 Observed Choices of Row Players in Constant-Sum 6x6

B. Procedures and Results of the Robust Estimation

We conduct a robust estimation procedure to check the robustness of the global maximum found by the DE algorithm and to show the difference in the local maxima between the two models in each dataset. We divide the reasonable range of each parameter into three intervals and randomly generate an initial value in each interval. For example, for parameter $\alpha \in [0,1]$ in the BLK* model, we randomly generate a relatively small initial value for α in the interval [0, 1/3], a relatively large initial value in the interval [2/3, 1], and a medium level initial value in the interval [1/3, 2/3]. For each parameter, we choose three levels of initial values, and thus, we can get 3^5 (243) different combinations of initial values for parameters.[23] We use the Matlab optimization toolbox to search local maxima by 243 times. Table A1 shows the brief results of this robust estimation.

Table A1 The brief results of the robust estimation

Datasets	DSG	nDSG	Zero-Sum	Constant-Sum	
	3x3	3x3	4x4	4x4	6x6
BLK*					
No. of local maxima	35	32	41	13	16
Global maximum (LL)	-871.11	-974.79	-6979.34	-1931.35	-2444.64
No. of combinations (converged to the global maximum)	10	90	19	75	91
No. of total combinations	243	243	243	243	243
No. of Obs.	1280	1280	5250	1600	1600
BCH					
No. of local maxima	6	6	13	24	12
Global maximum (LL)	-903.12	-969.61	-6918.06	-1932.62	-2407.42
No. of combinations (converged to the global maximum)	205	174	126	79	118
No. of total combinations	243	243	243	243	243
No. of Obs.	1280	1280	5250	1600	1600

The results of this procedure can provide two pieces of useful information. First, the estimation results confirm that the estimation result from the DE algorithm is the global

[23] To ensure comparability, we use the same combinations of initial values in the robust estimations of all five datasets. We also use the same initial values for λ' (λ''), ρ, γ_0, and γ_H between the BLK* and BCH models. For the initial values of λ in the BCH model, we simply derive the corresponding value of λ for each initial value of α, i.e., the proportion of level-0 players.

maximum in the robust estimation and the combination of estimates is unique in the global maximum in all five datasets. Second, we can compare the number of local maxima and the convergence of initial combinations between the two models. Note that the BCH model interprets the dynamic behavioral change only as changes in beliefs, but the BLK* model interprets that as either changes in beliefs or changes in reasoning levels. We conjecture that the BLK* model should have more local maxima than the BCH model, and meanwhile, the number of initial combinations converged to the global maximum should be smaller in the BLK* model than the BCH model. The results shown in Table A1 confirm our conjecture in the first three datasets, but in the latter two datasets of Constant-Sum games, the two models perform similarly in terms of the number of local maxima and the number of combinations converged to the global maximum.

C. Identification of Parameters in BLK* and BCH

In Appendix B, we have checked the parameters of the models are identified in the five datasets, i.e., no other combinations of parameter values fit equally well. However, for some other datasets, it is possible that the models cannot be identified. To check the severity of nonidentifiability, following Camerer and Ho (1999), we use the correlations between parameters based on the Jackknife estimations to detect possible identification problems of the BLK* and BCH models. For each parameter, we calculate the mean absolute correlation with other parameters in each dataset. Table A2 shows these statistics. We have not found the correlations between specific parameters are consistent in magnitude and sign across the datasets in both models. However, for the BLK* model, the mean absolute correlations are relatively high in the dataset of DSG. By contrast, for the BCH model, they are relatively high in the dataset of Constant-Sum 6x6. These results indicate that in some games, there are modest identification problems in both models.

Table A2 Mean absolute correlations of parameters

Datasets	DSG	nDSG	Zero-Sum	Constant-Sum	
	3x3	3x3	4x4	4x4	6x6
BLK*					
$\hat{\alpha}$	0.46	0.18	0.24	0.41	0.18
$\widehat{\lambda''}$	0.57	-	0.44	0.38	0.22
$\hat{\rho}$	0.47	-	-	-	-
$\widehat{\gamma_0}$	0.60	0.24	0.44	0.32	-
$\widehat{\gamma_H}$	0.61	0.13	0.20	0.50	0.11
BCH					
$\hat{\lambda}$	0.42	0.28	0.09	0.30	0.70
$\widehat{\lambda'}$	0.02	0.30	0.39	0.21	0.63
$\hat{\rho}$	-	0.36	0.42	0.29	0.79
$\widehat{\gamma_0}$	0.35	0.24	0.28	-	-
$\widehat{\gamma_H}$	0.41	0.18	0.56	0.43	0.60

D. Cross Predictions between Datasets

In Table A3, we use the estimates of Tables 3 and 5 from each model in each dataset (listed in the leftmost column) to predict the overall probability of a series of observed choices from each individual over periods in each dataset (listed in the first row). Then, we derive the average probability of these predicted probabilities across individuals for a single decision and report these average probabilities in Table A3.

As shown in all these panels, first, the estimates of each dataset predict the highest average probability in each dataset (i.e., italicized in each panel), which confirms that the estimates are of the global maximum of the log-likelihood function. Second, the estimates from analogous games predict second best in each dataset. For example, the estimates from nDSG predict second best in the dataset of DSG, and the estimates of Constant-Sum (CS) 6x6 predict second best in the dataset of CS 4x4. Third, however, the estimates from a very different game predict poorly in each dataset. For example, the estimates from CS 4x4 or CS 6x6 predict the lowest average probabilities in the dataset of DSG. For the models of SEWA, BCH, BWS, BWB, and SCH, these average probabilities are even lower than the prediction of random choices. These observations indicate that the estimated models discussed in the paper are falsifiable.

Table A3 Average probabilities of cross predictions between datasets

Panel A

Average probability of cross predictions (SEWA)

Datasets	DSG 3x3	nDSG 3x3	ZS 4x4	CS 4x4	CS 6x6
DSG 3x3	***47.88%***	45.25%	24.94%	25.40%	16.96%
nDSG 3x3	47.88%	***45.25%***	24.94%	25.40%	16.96%
ZS 4x4	34.52%	34.34%	***25.00%***	25.02%	16.68%
CS 4x4	12.39%	11.37%	22.54%	***26.37%***	17.63%
CS 6x6	13.00%	11.95%	22.66%	26.37%	***17.63%***
Random choice	1/3	1/3	1/4	1/4	1/6

Panel B

Average probability of cross predictions (BLK)*

Datasets	DSG 3x3	nDSG 3x3	ZS 4x4	CS 4x4	CS 6x6
DSG 3x3	***50.62%***	44.48%	24.84%	25.43%	16.86%
nDSG 3x3	48.43%	***46.69%***	25.15%	25.51%	17.14%
ZS 4x4	40.39%	37.53%	***26.46%***	27.54%	18.56%
CS 4x4	47.29%	41.71%	26.05%	***29.91%***	21.18%
CS 6x6	33.61%	29.63%	25.72%	29.54%	***21.70%***
Random choice	1/3	1/3	1/4	1/4	1/6

Panel C

Average probability of cross predictions (BCH)

Datasets	DSG 3x3	nDSG 3x3	ZS 4x4	CS 4x4	CS 6x6
DSG 3x3	***49.38%***	46.39%	25.08%	25.49%	17.15%
nDSG 3x3	49.08%	***46.88%***	25.19%	25.54%	17.16%
ZS 4x4	31.33%	28.51%	***26.77%***	27.42%	18.40%
CS 4x4	27.34%	27.18%	24.42%	***29.88%***	22.05%
CS 6x6	27.93%	27.75%	24.52%	29.68%	***22.21%***
Random choice	1/3	1/3	1/4	1/4	1/6

Panel D

Average probability of cross predictions (BWS)

Datasets	DSG 3x3	nDSG 3x3	ZS 4x4	CS 4x4	CS 6x6
DSG 3x3	***43.11%***	29.31%	25.55%	28.74%	19.96%
nDSG 3x3	33.45%	***34.98%***	25.24%	25.36%	16.88%
ZS 4x4	27.94%	27.90%	***26.57%***	27.55%	18.48%
CS 4x4	27.34%	27.18%	24.42%	***29.88%***	22.05%
CS 6x6	27.91%	27.75%	24.52%	29.68%	***22.21%***
Random choice	1/3	1/3	1/4	1/4	1/6

Panel E

Average probability of cross predictions (BWB)

Datasets	DSG 3x3	nDSG 3x3	ZS 4x4	CS 4x4	CS 6x6
DSG 3x3	***49.37%***	46.31%	25.07%	25.48%	17.15%
nDSG 3x3	49.08%	***46.88%***	25.19%	25.54%	17.16%
ZS 4x4	30.88%	27.74%	***26.77%***	27.38%	18.37%
CS 4x4	26.81%	26.81%	24.91%	***29.74%***	21.80%
CS 6x6	29.43%	29.61%	25.16%	29.42%	***21.99%***
Random choice	1/3	1/3	1/4	1/4	1/6

Panel F

Average probability of cross predictions (SCH)

Datasets	DSG 3x3	nDSG 3x3	ZS 4x4	CS 4x4	CS 6x6
DSG 3x3	***43.03%***	29.73%	25.36%	28.92%	20.29%
nDSG 3x3	33.72%	***34.83%***	25.23%	25.32%	16.85%
ZS 4x4	28.16%	27.94%	***26.57%***	27.52%	18.46%
CS 4x4	26.81%	26.81%	24.91%	***29.74%***	21.80%
CS 6x6	29.43%	29.61%	25.16%	29.42%	***21.99%***
Random choice	1/3	1/3	1/4	1/4	1/6