Learning Games

Nobuyuki Hanaki*,a,b, Ryuichiro Ishikawa[c], Eizo Akiyama[c]

[a] *GREQAM and Université de la Méditerranée, 2 Rue de la Charite, Marseille, 13002, France*
[b] *Department of Economics, Graduate School of Humanities and Social Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki. 305-8571, Japan*
[c] *Department of Social Systems and Management, Graduate School of Systems and Information Engineering, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki. 305-8573, JAPAN*

Abstract

This paper presents a model of learning about a game. Players initially have little knowledge about the game. Through playing the same game repeatedly, each player not only learns which action to choose but also constructs a personal view of the game. The model is studied using a hybrid payoff matrix of the prisoner's dilemma and coordination games. Results of computer simulations show that (1) when all the players are slow at learning the game, they have only a partial understanding of the game, but might enjoy higher payoffs than in cases with full or no understanding of the game; (2) when one player is quick in learning the game, that player obtains a higher payoff than the others. However, all can receive lower payoffs than in the case in which all players are slow learners.

Key words: Learning, Subjective views, Computer simulation, Inductive game theory
JEL code: C72, D83

*Corresponding author
Email addresses: hanaki@dpipe.tsukuba.ac.jp (Nobuyuki Hanaki), ishikawa@sk.tsukuba.ac.jp (Ryuichiro Ishikawa), eizo@sk.tsukuba.ac.jp (Eizo Akiyama)

1. Introduction

In standard game theory, players are assumed to have well-formed beliefs and knowledge of the structure of the game they play. The origins of their beliefs and knowledge are rarely studied. The validity of this assumption is questioned, however, in experimental economics.[1] According to Camerer (2003, p. 474), "what game do people think they are playing?" is a top 10 most important open research question in experimental game theory.

If players do not understand a game completely, or misunderstand it, how do they learn about the true game? In the relevant literature, both theoretical and experimental, on learning in games, this question has been seldom addressed. The literature has mainly addressed learning about how to play a game rather than on learning about the game itself.[2] An exception in experimental game theory is Oechssler and Schipper (2003). They conducted a set of experiments in which subjects did not know the payoffs of their opponent and were given incentive to learn about them in 2×2 games. The authors constructed the games which subjects perceived they were playing—the subjective games—from the data. They found that the subjective games differed frequently from the games that were actually being played.

Independently from the developments in experiments, Kaneko and Kline (2007, 2008) initiated a new theory called "inductive game theory" to shed a light on how players gain understandings of a game, and to study implications of players having only a partial understanding of it. For example, Kaneko and Kline assume that players have little knowledge of the game they play. By playing the game repeatedly, the players obtain experiences and accumulate these memories. Based on the memories, they inductively

[1] Camerer (2003, p. 474) gives an example of a student who participated in an experiment at Caltech. The students confused the coordination game used in the experiment with the prisoner's dilemma game, and "defected" continuously.

[2] Fudenberg and Levine (1998) provided a detailed survey of the theoretical literature of learning in games. See, for example, Crawford (1995), Cheung and Friedman (1997), Mookherjee and Sopher (1997), Erev and Roth (1998), and Camerer and Ho (1999) for the experimental learning literature. Arifovic et al. (2006) is an interesting exercise that is intended to compare the various learning models in terms of their abilities to replicate human behaviors observed in laboratory experiments.

form their personal views about the game. Following this line, Akiyama et al. (2008) undertakes a simulation study of a player 's learning about the structure of a game. They consider a simple one-person example, and simulate the process by which a player experiences and accumulates memories about the game structure.

Following postulates of behavior and memories in Kaneko and Kline (2008, Section 3.2, p.1343), Akiyama et al. (2008), in the simulation, introduces two types of memories, *short-term* and *long-term* memories as a player 's cognitive ability. They study transitions form short-term memories to long-term ones in response to the frequency of experiences and show the difficulty of learning about the structure within practical finite time spans. While this paper is also motivated to investigate inductive learning, our focus is rather dynamics of players ' learning when setting a probabilistic behavior model as a variant on informal theory in Kaneko and Kline (2008).

In this paper, we describe a model in which players play a normal form game repeatedly and learn not only about how to play the game but also about the game itself. A normal form game consists of the set of players, the set of available actions (or strategies), and the payoff function for each player. Therefore, learning about a game means that players do not know some of these components and learn about them. As described in this paper, we assume that a player knows about the set of actions available to himself and the number of opponents. However, initially the player does not know about the set of actions available to his opponents or anyone's payoff functions. The player learns about them—in particular, his own payoffs associated with possible outcomes—by playing the game repeatedly. Therefore, for the discussion presented herein, players are learning about different aspects of a game than those that were examined in Oechssler and Schipper (2003).

To model how players build their personal views about a game from playing it repeatedly, we have adopted informal theory in inductive game theory (Kaneko and Kline (2007, 2008); Akiyama et al. (2008)), which is related with a concept in cognitive science: The role that *autobiographical memory* plays in learning from everyday life events

(Linton, 1982; Wagenaar, 1986). An autobiographical memory is a memory of frequently repeated events. It is an abstract script so that the details, such as the date of occurrence, are lost; only the general facts about the events remain. To replicate such a mode of learning about events in our model, players are assumed to have memory of two types, *short-term* and *long-term memory*.[3]

Short-term memories are temporary memories of outcomes of playing the game, i.e., the actions chosen by players and the payoffs received by the player.[4] A short-term memory remains in the player's mind only for a certain number of periods, and vanishes thereafter. The outcome will be retained as a long-term memory if the same outcome is repeated frequently enough while the short-term memories of it remain in the player's mind. In other words, an outcome of the game will be retained in the player's mind as a long-term memory if it has been recalled frequently during a specified period. A long-term memory permanently associates, in the mind of the player, an outcome of the game with a payoff. Once an outcome of the game is engraved in the player's mind as a long-term memory, it remains there forever. Then we say that the player has learned the part of the game corresponding to it. The personal view of a player about the game is simply the part of the game that the player has learned.

In addition to learning about the payoffs, players learn which action to choose. The latter is modeled based on the reinforcement learning model. When a player does not know any of the payoffs, only the realized payoffs will be used. As the player learns some parts of the game, she starts to infer what the payoffs could have been if she had acted differently—at least for the parts of the game she knows. Therefore, learning about performance of actions will be based not only on the realized payoffs but also on the forgone payoffs where possible. We have studied this model in a 3×3 game that embeds both a prisoner's dilemma and a coordination game. Through a series of

[3]The model of the mind's memory system composed of short-term and long-term memory was first proposed by James (1890) and established by Atkinson and Shiffrin (1968). Long-term memory is classifiable into episodic memory and semantic memory (Tulving, 1972). Autobiographical memory is a type of episodic memory for information related to *oneself* (Brewer, 1986).

[4]In general, the word "memory" can mean either a store of information or information itself. As described in this paper, we mainly use "memory" in the latter sense.

computer simulations of the model, we demonstrate what kind of personal views the players tend to form and what kind of behavior emerges when players' personal views and their behavior coevolve.

Results show that, when all the players are very slow in learning about a game, they will only have a limited understanding (a partial view) of the game. Such a limited understanding, however, can be beneficial for them. They might enjoy higher payoffs than in the case of full or no understanding of the game. The players enjoy high payoffs. Therefore, their behaviors do not change. Consequently, neither do their personal views. Therefore, their views remain partial. When one of the players is quick in learning about a game, he can obtain a higher payoff than the other players who are slow in learning. However, in this case, all the players, even the fast-learning player, might obtain lower payoffs than in the case where all the players are very slow learners.

The remainder of the paper is organized as follows. Section 2 presents the model in detail. Results of the model simulation are summarized in Section 3. Section 4 offers an explanation of the results, and Section 5 concludes this paper.

2. A Model of Game Learning

We consider a two-person game. The set of players is $\{1, 2\}$, and each player $i \in \{1, 2\}$ has a set S^i of available actions. Initially, each player only knows the set of actions that is available to him, but he does not know the action sets of the others nor the payoffs, i.e., he does not know the game he is facing. Through playing the game repeatedly, players not only learn which action will bring about higher payoffs[5] but also form their views of the game they are playing. Below, we first discuss how we model the formation

[5]Because players are repeatedly playing the game, in principle, it is possible for them to employ repeated game strategies and be learning about performance of such strategies instead of stage game actions. See Hanaki et al. (2005) for a model in which players are learning about which repeated game strategies to use. In this paper, however, we do not consider such possibilities. In the situation we are considering, players initially do not know the set of available actions for the opponents and learn about them as they play the game. It means that a player learns about the set of repeated game strategies available for him as he plays the game. This expansion of the strategies set is an interesting problem to consider in itself: we therefore leave this for future research.

of personal views by players, and the representation of such views. Once we define a player's personal view of the game, we formulate how the player learns which action to choose based on the past outcomes and his view of the game.

2.1. Formation of personal views

We assume that a player has memories of two types, *short-term* and *long-term memories*, like Akiyama et al. (2008). A short-term memory is a memory of an outcome of playing the game, i.e., the actions chosen by players and the resultant payoff received by the players. The short-term memory remains in the player's mind only for a certain number of periods, and vanishes thereafter. The outcome will be retained as a long-term memory if the same outcome is repeated frequently enough while the short-term memories of it remain in the player's mind. Once outcomes of the game are engraved in the player's mind as a long-term memory, that memory remains there forever. If that occurs, we say that the player has learned the part of the game corresponding to it.

More precisely, player i is characterized by his short-term memory length m^i and cognition threshold $k^i (\leq m^i)$. The short-term memory length is the number of periods before a short-term memory vanishes from his mind. The cognition threshold represents the number of repetitions needed for short term memories of an outcome to be retained as a long-term memory. Because short-term memories vanish after m^i periods, an outcome (s^i, s^j) will be retained in i's mind as a long-term memory if an outcome has been recalled k^i times in the m^i most recent interactions. Once outcome (s^i, s^j) is recorded as a long-term memory, the player knows the payoff which can be received if the outcome is realized in the future. The transformation of an outcome of the game in the mind of a player from a short-term memory to a long-term memory in this paper plays a similar role as in the *autobiographical memory* in cognitive science. Autobiographical memory is the memory of everyday life events. As shown by Linton (1982) and Wagenaar (1986), this memory retains a repeated event as an abstract script. That is, when keeping such repeated events, details such as date of occurrence are lost, and only the general facts

about the events remain. In addition, if an event is not repeated, the memory of the event will not remain as an autobiographical memory.

We define a player's personal view of the game based both on the objective payoff matrix and the set of long-term memory in the player's mind. The set of long-term memory in the player's mind might change over time. Therefore, the player's personal view of the game will also change.

Let Π be the objective payoff matrix of the game under consideration, and let Π^i represent the part of the payoff matrix that corresponds to what player i receives. Namely, in a two person game,

$$\Pi^i = \begin{pmatrix} \pi^i(s_1^i, s_1^j) & \dots & \pi^i(s_1^i, s_{n^j}^j) \\ \vdots & \ddots & \vdots \\ \pi^i(s_{n^i}^i, s_1^j) & \dots & \pi^i(s_{n^i}^i, s_{n^j}^j) \end{pmatrix},$$

where n^i and n^j are the numbers of actions in S^i and S^j, respectively, and $\pi^i : S^i \times S^j \to \mathbf{R}$ is player i's payoff function. We assume that $\pi^i(s^i, s^j) \neq 0$ for all $s^i \in S^i, s^j \in S^j, i, j \in \{1, 2\}$ with $j \neq i$ because we assign a special meaning to value zero in the subjective payoff matrix as defined below.

Let $L^i(t)$ be the matrix that represents the state of the long-term memory in the mind of player i at period t, where each element of the matrix takes value zero or one; $L^i_{s^i,s^j}(t) \in \{0, 1\}$. $L^i_{s^i,s^j}(t)$ takes value zero when outcome $\left(s^i, s^j\right)$ is not in player i's mind as a long-term memory at period t, and it is one otherwise. We assume that, initially, players do not know about any of the outcomes; that is, $L^i_{s^i,s^j}(0) = 0$ for all $(s^i, s^j) \in S^i \times S^j$.

The personal view of the game for player i at period t, $\widetilde{\Pi}^i(t)$, is definable as

$$\widetilde{\Pi}^i(t) = L^i(t) \cdot \Pi^i. \tag{1}$$

Therefore, $\widetilde{\Pi}^i_{s^i,s^j}(t)$ is zero when player i has not learned of the outcome $\left(s^i, s^j\right)$ at period

t, and it is equal to $\pi^i(s^i, s^j)$ otherwise. We designate this matrix the subjective payoff matrix for player i at period t. Now we proceed to a discussion of how players learn which action to choose.

2.2. Learning about performance of actions

We assume that a player's recent experiences from choosing (as well as not choosing) an action are summarized by his "attraction" for the action. In each period, players choose their actions based on their attractions for each action. It is through the evolution of attractions that players learn.

Let $A_s^i(t)$ denote player i's attraction for action $s \in S^i$ at period t. The probability that player i chooses action s at period t, $p_s^i(t)$, depends on the player's attraction as follows:

$$p_s^i(t) = \frac{e^{\lambda^i A_s^i(t)}}{\sum_{k \in S^i} e^{\lambda^i A_k^i(t)}}. \tag{2}$$

Parameter λ^i in the logistic transformation represents the extent to which actions with higher attractions are favored in action choice. When $\lambda^i = 0$, all actions are equally likely to be chosen irrespective of their attraction. As λ^i becomes larger, actions with higher attractions become disproportionally more likely to be chosen. In the limiting case where $\lambda^i \to \infty$, the action with the highest attraction is chosen with probability one. The logistic transformation introduced here is common in the literature on learning in games as well as experimental game theory to model the action choices of subjects in laboratory experiments (see, for example, McKelvey and Palfrey, 1995; Erev and Roth, 1998; Camerer, 2003). Other applications of such a logistic choice model include, for example, Brock and Hommes (1997, 1998), in which agents learn and choose among various price forecasting strategies in market settings. We assume that all the actions have the same attraction for all the players at the beginning of the game, i.e., $A_s^i(0) = 0$ for all i and $s \in S^i$. Therefore, initially, all the actions are equally likely to be chosen irrespective of λ^i.

Attractions evolve as

$$A_s^i(t+1) = \frac{1}{h^i} \sum_{\tau=0}^{h^i-1} R_s^i(t-\tau), \tag{3}$$

where $h^i = \min(m^i, t+1)$.[6] $R_s^i(t)$ is a stimulus the player receives for action s at period t, which depends on the outcome of the game as well as the player's understanding of the game in period t in the following manner:

$$R_s^i(t) = \begin{cases} \pi^i(s^i(t), s^j(t)) & \text{if } s = s^i(t) \\ \widetilde{\Pi}^i_{s,s^j(t)}(t) & \text{otherwise,} \end{cases} \tag{4}$$

where $s^i(t)$ represents the action chosen by player i in period t. Equation (4) states that the stimulus player i receives for action $s \in S^i$ at t is the realized payoff when he chooses s at period t, i.e., $s = s^i(t)$, irrespective of the status of long-term memory in i's mind. If player i does not choose s at period t, the stimulus follows i's subjective payoff matrix. Therefore, if the payoff consequence of $(s, s^j(t))$ is in i's mind as a long-term memory, then the stimulus for action s will be the forgone payoff, i.e., the payoff player i could have obtained if he had chosen action s in period t, given the action chosen by the opponent in that period, $s^j(t)$. Otherwise, the stimulus will be zero.

In this definition of stimulus, we are assuming that once a player understands some of the payoffs of the game, the player can infer what the payoffs could have been if he had acted differently for the part of the game he knows. For the parts of the game he does not know, he cannot make such inferences.

The proposed model of learning about performance of actions builds on two models of action learning commonly studied in the literature: learning based only on realized payoffs (see, e.g., Erev and Roth, 1998) and learning based on both forgone and realized payoffs (see, e.g., Camerer and Ho, 1999). The bridge between these two in our model is the long-term memory or personal views. Indeed, when $L^i_{s^i,s^j}(t) = 0$ for all t, $(s^i, s^j) \in$

[6] $h^i = \min(m^i, t+1)$ is to take care of the early periods so that the game has not been played m^i times.

$S^i \times S^j$, players in our model learn about performance of actions based only on realized payoffs (RL model in below). In contrast, when $L^i_{s^i,s^j}(t) = 1$ for all t, $(s^i, s^j) \in S^i \times S^j$, players always learn based on both realized and forgone payoffs (FP model in below).[7] It is the dynamics of long-term memory that make our model different from existing learning models. In the next section, we describe results of computational simulation of our model.

3. Simulation Results

Because one focus of this study is to investigate what personal views emerge from our model, in this paper, we consider symmetric 3×3 games which embed two symmetric 2×2 games.[8] The two games embedded are prisoner's dilemma and coordination game. As players learn the game, they may see themselves facing a prisoner's dilemma-type situation, a coordination game-type situation, or something else. Depending on how players understand the game, their behaviors might vary.[9]

Each player $i \in \{1, 2\}$ has three available actions $\{s^i_1, s^i_2, s^i_3\}$. The objective payoff matrix Π with a parameter $a \in (0, 0.5)$ is given as follows.

	s^2_1	s^2_2	s^2_3
s^1_1	$1-a$, $1-a$	0, 1	1, 0
s^1_2	1, 0	a, a	a, 0
s^1_3	0, 1	0, a	$1-a$, $1-a$

Four cells in the upper left corner correspond to a prisoner's dilemma game and four cells in the lower right corner are a coordination game.

[7] As shown by Camerer and Ho (1999), when the learning is based both on realized and forgone payoffs, the behavior the model generates will be equivalent to that generated by *fictitious play* with probabilistic action choice, where each player, knowing the whole payoff structure, chooses the best response, although probabilistically, to the empirical frequency of choice (action) of his opponent.

[8] 3×3 games are the minimal symmetric games that can involve partial strategic structures where players have choices.

[9] We have also considered these two 2×2 games separately. The results are discussed briefly in Appendix C.

As described in the previous section, we have assigned a special meaning to the zero in the subjective payoff matrix, the payoff consequences of outcomes that are not retained as long-term memories. In order not to have zero in the objective payoff matrix, we added $b = 0.01$ to all the payoffs.[10] b is not shown in the above payoff matrix for clarity of exposition. The unique pure Nash equilibrium of this game is (s_2^1, s_2^2), with payoff (a, a).

Parameter a reflects the severity of the dilemma in the prisoner's dilemma game as well as the risk–payoff tradeoff in the coordination game. In the prisoner's dilemma game embedded here, the lower a is, the larger the aggregate loss of not choosing (s_1^1, s_1^2) becomes. In the embedded coordination game, if $1/3 < a < 0.5$, (s_2^1, s_2^2) is the risk-dominant equilibrium whereas (s_3^1, s_3^2) is the payoff-dominant equilibrium. It is interesting to see what kind of views players construct and what kind of behavior they learn over time under various values of a.

In the simulation analysis described below, we first specifically examine the cases in which all players have the same short-term memory length $m^i = m$ and the same cognition threshold $k^i = k$. We then proceed to cases where players have the same short-term memory length but different cognition thresholds. Throughout the paper, we assume that the sensitivity of action choices to the attractions are the same across players, $\lambda^i = \lambda$ for all $i \in \{1, 2\}$.

3.1. Case of identical short-term memory length and threshold

We first consider the case in which the short-term memory length of the players and their cognition thresholds are identical across players. In such cases, we have four parameters in our model: a defines the payoff matrix, m and k determine the length of the players' short-term memories and their cognition threshold, and λ governs the importance of attractions in action choices. We will present results based on a particular set of parameter values ($m = 5$, $\lambda = 5.0$, $a = 0.25$ while varying k). The dependencies of the results on the parameter values are discussed in the appendix.

[10] The results remain the same if we subtract $b = 0.01$ from all the payoffs.

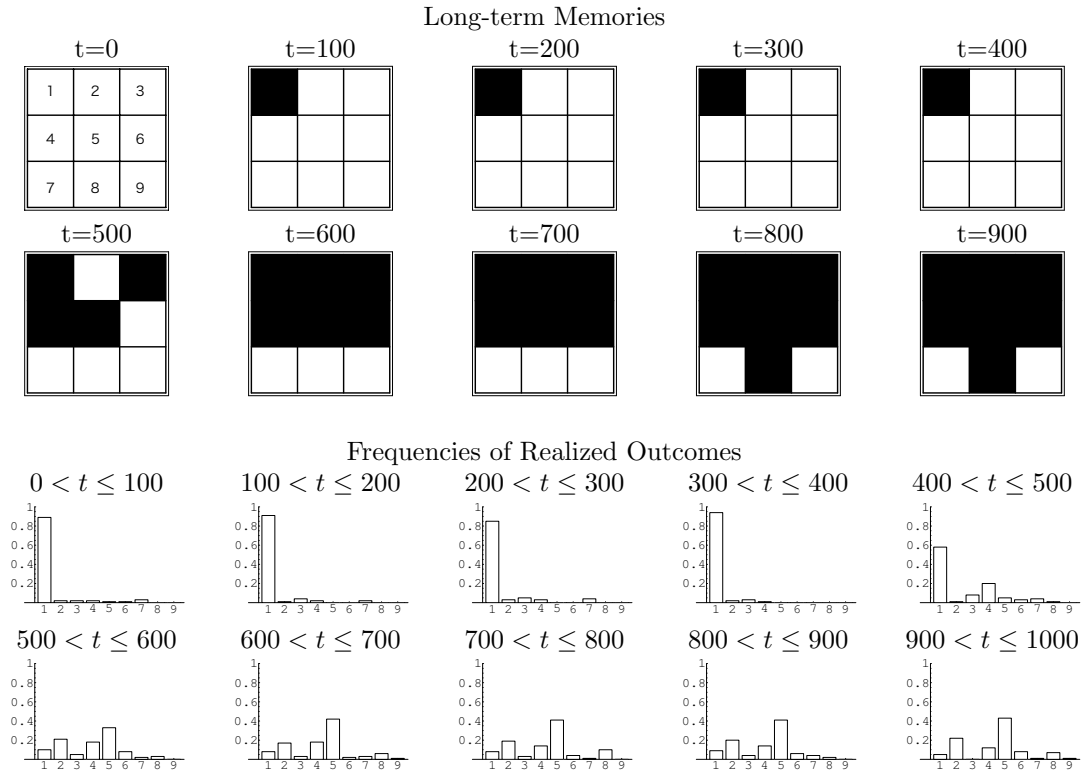

Figure 1: Evolution of the long-term memory (top) and the corresponding frequencies of realized outcomes (bottom) from a single simulation run. $m = 5$, $k = 3$, $\lambda = 5.0$, $a = 0.25$. The black cells represent those outcomes recorded as a long-term memory.

Figure 1 shows an outcome of a simulation run.[11] The parameters are set so that $m = 5$, $k = 3$, $\lambda = 5.0$, and $a = 0.25$. To portray the evolution of players' respective long-term memories (top panel), the outcomes that are retained as long-term memories at a given point in time are represented by the black cells. Nine cells in the objective payoff matrix are numbered 1–9, as shown in the status of long-term memories at $t = 0$. The frequencies of realization of each outcome of the game are represented by the height of the corresponding bars in the bottom panel. In this simulation run, outcome 1 has been realized with a high frequency in earlier periods. As a result, outcome 1 became a long-term memory of the two players. Between periods 400 and 500, however, because of a player's deviation from playing action 1, the players learned other outcomes. Consequently, the behaviors of the players change quite drastically in the later periods. As the figure shows, in the later period, outcome 5, which is the Nash equilibrium of the game, is realized with the highest frequency.

The result portrayed in Fig. 1 is merely an example of how the players' understanding of the game and their behaviors coevolve as players repeatedly play the game. However, it is not the representative result. In fact, there can be many other patterns of coevolution. Instead of enumerating all the possible results, we specifically examine averaged results[12] below.

Figure 2 presents the results of simulations for $a = 0.25$. In each figure, the average payoffs of the row-player over time (top),[13] the average status of long-term memories at $t = 500$ (bottom), and the average frequencies of realized outcomes for $500 \leq t \leq 1000$ (middle) are shown for $k = 1$ (left), $k = 3$ (center), and $k = 5$ (right). Except for the average status of the long-term memories, the outcome of the two other models of learning, the one based only on realized payoff (RL model) and the other based on both realized and forgone payoff (FP model) are reported. In fact, as discussed in the last

[11] A single simulation run consists of 1000 interactions by a pair of players. One period in the simulation corresponds to one interaction by the pair of players.

[12] For each set of parameter values, we take the average of the results generated by 100 simulation runs, while giving varying random seeds for each run.

[13] The average payoff of the column-player closely resembles that of the row-player.

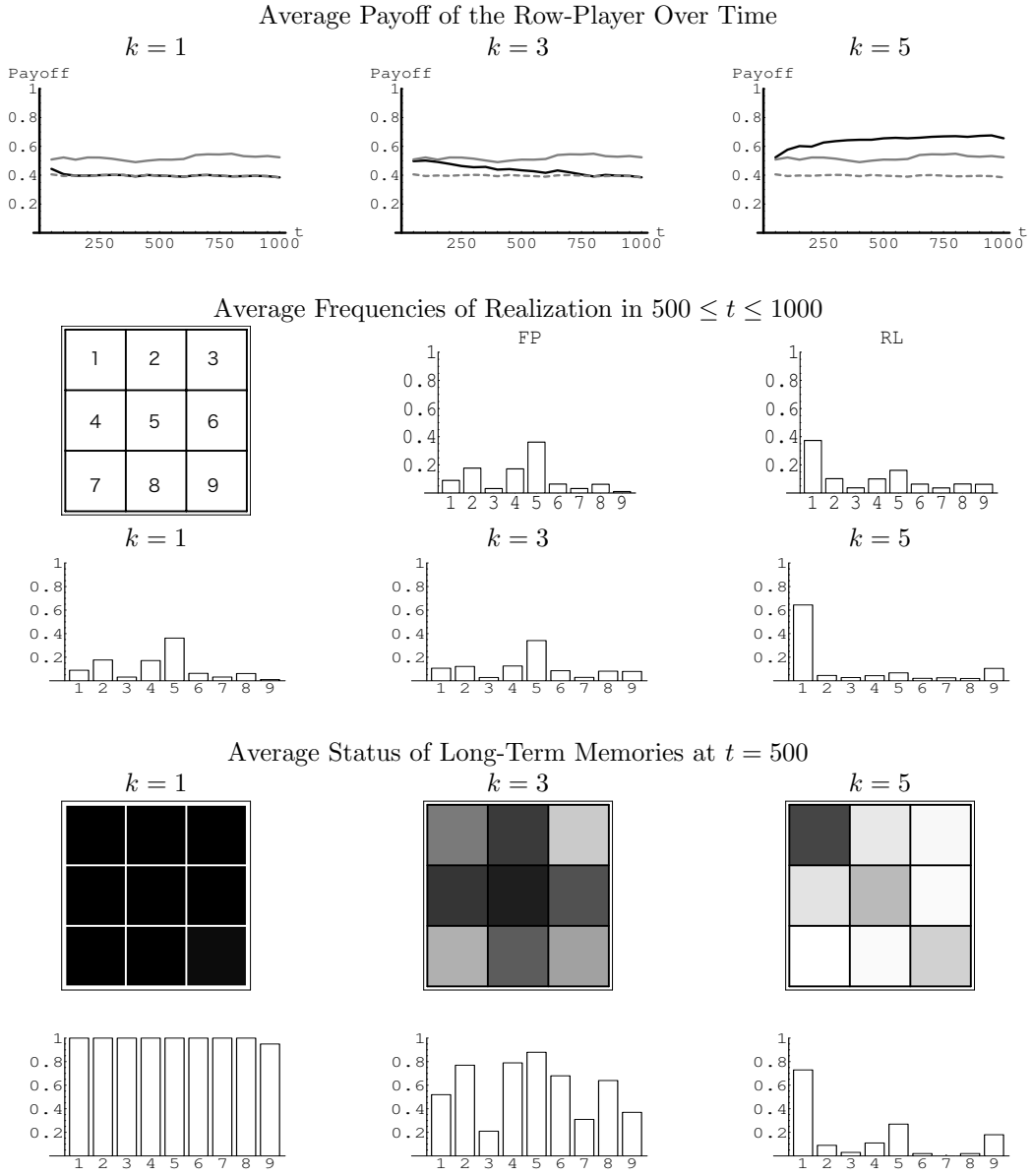

Figure 2: Average payoff of the row-player over time (top), average frequencies of realized outcomes (middle), and average status of long-term memories (bottom) for various k. $m = 5$, $a = 0.25$, $\lambda = 5.0$. For the average payoff, the result of our model is presented in solid black, the solid gray represents the RL model, and the dashed gray represents the FP model. For the average status of the long-term memories, the darker gray corresponds to the higher likelihood that the outcome is retained as a long-term memory, which is also represented by the height of bars.

paragraphs of the model section above, these two additional models are two special cases of our model. Namely, the RL model is the case in which agents never learn about the game, i.e., $L^i_{s^i,s^j}(t) = 0$ for all t, $(s^i, s^j) \in S^i \times S^j$. The FP model is the case in which agents know all the payoffs from the beginning, i.e., $L^i_{s^i,s^j}(t) = 1$ for all t, $(s^i, s^j) \in S^i \times S^j$.

In the top panel of the figure, the outcomes from the three models are shown. The average payoff of the row-player from the RL model ("RL players") is shown with solid gray. That from the FP model ("FP players") is shown with dashed gray. The result of our model (game learning model, "GL players") is shown in solid black. The middle panel shows, according to the height of the bars, the frequencies with which outcomes corresponding to cells numbered from 1 to 9 are realized. The FP in the figure stands for the outcome of the "FP model" whereas RL stands for that of the "RL model". In the bottom panel, the darkness of a cell and the corresponding height of the bars show the proportion of the simulation runs for which the outcome was recorded as a long-term memory at $t = 500$.

When $k = 1$, one can see that the payoff received by GL players quickly converges to those received by FP players. When $k = 1$, players learn all the payoffs of the game quite quickly. As one can see in the figure, by $t = 500$, all the payoffs are known by the players almost all the time.[14] Once GL players learn of all the payoffs, their behaviors become equivalent to those of FP players.[15]

The convergence in the average payoff is much slower in the case of $k = 3$. In this case, players do not learn all the payoffs as in the $k = 1$ case. Players learn, however, the outcomes 2, 4, 5, 6, and 8 by $t = 500$ in the most simulation runs. The result shows that such a partial understanding is sufficient for players to choose the Nash equilibrium action with a high probability.

[14] In fact, when $k = 1$, players understand all the payoffs by period 100 in most simulations.

[15] It is an interesting outcome that RL players play cooperative action (action 1) more often and receive higher average payoffs than FP players. Similar outcomes, i.e., players learn to cooperate in the setting where no punishment mechanism or explicit communication exists between agents, is reported by Waltman and Kaymak (2008) for repeated Cournot oligopoly games in which agents learn based on Q-learning, which is a reinforcement learning model used in the field of artificial intelligence.

Do partial understandings of the game always lead players to choose the Nash equilibrium action with a higher likelihood? The results from the $k = 5$ case shows that the answer is no. When $k = 5$, GL players receive higher payoffs than both RL and FP players. The partial understandings of the game that caused this case give more benefits to players than full or no understanding of the game. The figure shows that players learn outcome 1, which Pareto dominates Nash equilibrium outcome 5, in most simulation runs. Although outcomes 5 and 9 are also learned, such cases are infrequent.

The result arises from both a partial view of the game and the players' limited ability to acquire the view. As described in Section 2, it is hard for GL players with high cognition thresholds k to retain the outcomes as a long-term memory. Outcome 1 realized by these players is stable once it becomes a long-term memory because it is hard for player 1 (player 2) to encounter outcome 4 (2) sufficiently many times to learn that it is more attractive than outcome 1, which is already engraved in his mind as a long-term memory.

It is noteworthy that players' action choices and their understandings of the game coevolve in our model. Therefore, not only do players benefit from their limited understanding of the game, but also because they benefit from such a limited understanding, their behaviors do not change. Therefore, their views remain partial.[16]

This result is quite interesting and illuminates the possibility that, because we live in a very complex society, it might not be feasible for us to learn the true or complete interactive environment that we face. Our understanding of the environment might be very limited, but as long as we are satisfied with the outcomes, we do not actively try to learn the true environment (or do not try and see what will happen if we do something different from what we normally do). Therefore, our understandings remain limited. Of course, it is quite possible that because of our limited understanding and the lack of exploration, we are not receiving a higher payoff, which could be obtainable if we really understood the complete environment.[17]

[16]The dependency of the results on parameter values is discussed in Appendix A, which shows that the result holds in quite a large parameter space as long as players' cognition thresholds k are sufficiently high, i.e., close to their short-term memory length m.

[17]One might wonder whether 1000 periods is sufficiently long. If we consider a very long simulation,

3.2. Case of identical short-term memory length but different thresholds

In the previous subsection, we described the case in which players have identical short-term memory length m and cognition threshold k. When all the players are slow in constructing their personal view of the game (k close to m), they can obtain a higher payoff than when they are quick in learning the game structure (small k). What happens if two players who have the same short-term memory length $m^1 = m^2 = m$, but different cognition thresholds, interact? We consider such cases in this subsection.

Figure 3 presents typical dynamics of long-term memories of two players and the frequencies of realized outcomes when player 1's cognitive threshold is 1, $k^1 = 1$, and that of player 2 is 5, $k^2 = 5$. Both the players have the same short-term memory length, $m = 5$. The sensitivity of action choices to attractions, λ, are set equal to 5, and the payoff matrix is such that $a = 0.25$.

In this particular simulation run, both players learn outcomes 1 and 5 by period 100. Although player 2 only learns these two, player 1 learns all the outcomes except for outcome 9. By period 500, player 1 learns all the outcomes, but player 2's understanding remains limited to outcomes 1, 4, and 5.

The dynamics of the frequencies of the realized outcomes are quite interesting. After the first 100 periods, it was outcome 1 that had been realized the most. Beyond period 100, outcomes 4 and 5 are realized with higher frequencies than outcome 1. It is noteworthy that because player 1 knows almost all the payoffs, when both players are happily choosing action 1 (therefore outcome 1 is realized), player 1 can infer that if he chooses action 2 while player 2 continues choosing action 1, he can get a higher payoff (associated with outcome 4). Therefore, there is a high chance that he will change his behavior. However, if player 1 indeed chooses action 2, player 2 might think that it is

GL players can learn more about the game because of the probabilistic action choice. Therefore, we can say even without computer simulation that eventually, in a very long run, the GL players learn the entire game and behave exactly the same as FP players. This paper presents an investigation of the result of 1000-period computer simulations because our interest is not necessarily in the convergent states but in the construction of personal view by the players with limited abilities. Of course, how long it takes for GL players to learn the entire game depends on the model parameters. See Appendix B for more discussion.

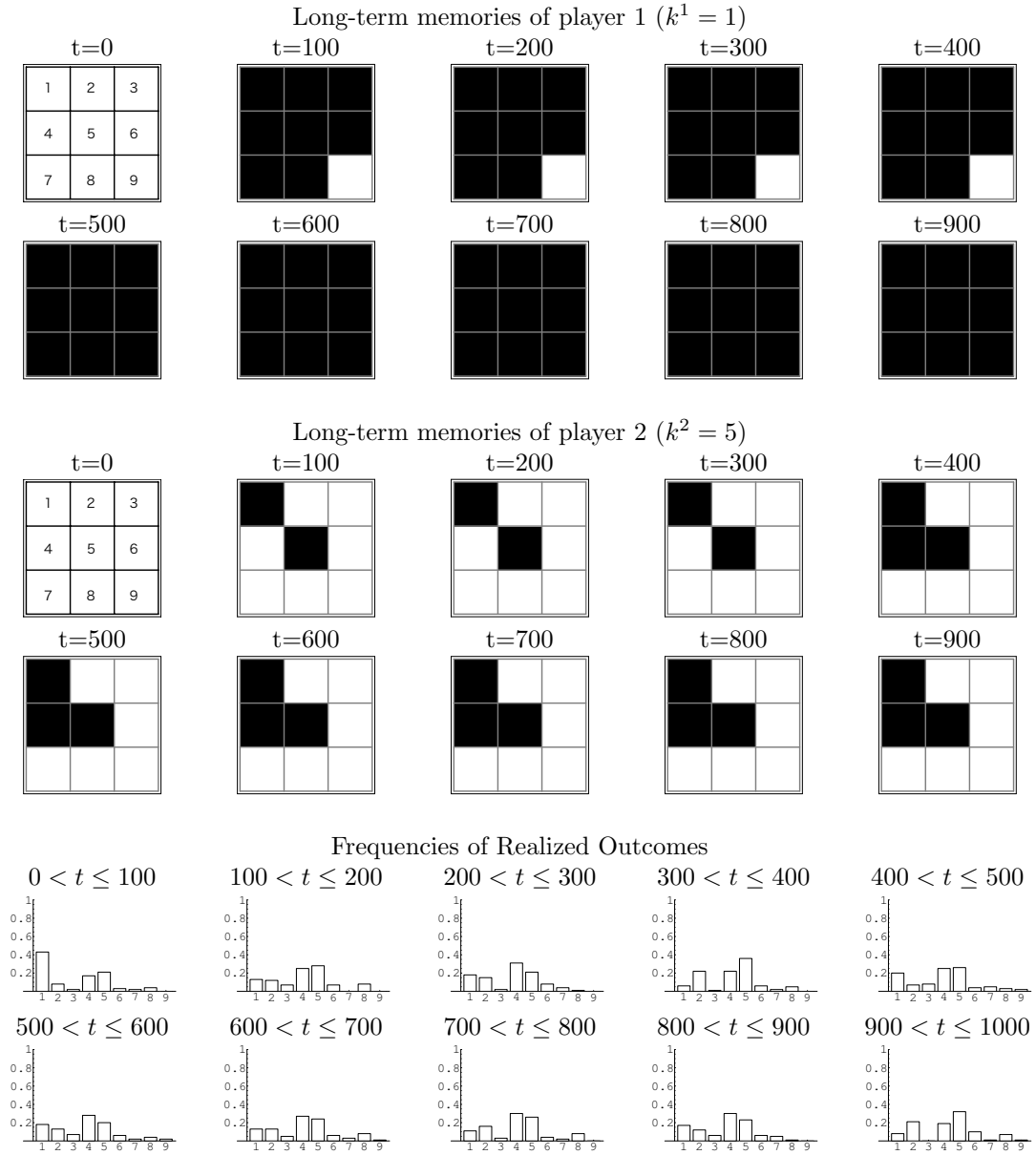

Figure 3: Evolution of the long-term memory for a player with low k, $k^1 = 1$, (top) and high k, $k^2 = 5$ (middle) and the corresponding frequencies of realized outcomes (bottom) from a single simulation run. $m = 5$, $\lambda = 5.0$, $a = 0.25$. The black cells represent the outcomes that are retained as long-term memories.

better to choose action 2 instead of action 1. (Recall that player 2 knows the payoff associated with outcome 5!) Such learning will result in both players indeed choosing action 2: the Nash equilibrium.

This example shows that even if only one of the players is quick in learning about the game, two players might learn to choose the actions that correspond to the Nash equilibrium outcome. It is not necessary that both the players be quick in learning about the game to reach the Nash equilibrium.

Figure 4 presents the averaged results[18] of the simulation runs. It shows, for three different thresholds ($k = 1, 3, 5$) of player 1, average payoffs of players over time (top), average frequencies of realized outcomes (middle), and average status of long-term memories for two players (bottom). In these simulation runs, player 2's thresholds are fixed at $k^2 = 5$. Other parameter values are $m^1 = m^2 = 5$, $a = 0.25$, and $\lambda = 5.0$. We are presenting the $k^1 = k^2 = 5$ case as the benchmark.

In the top panel, the average payoff of the player with a lower k, player 1, is the solid line, whereas that of the player with a high k, player 2, is the dashed line. In the figure, player 1 receives a higher payoff than player 2 because a player who is quick in learning about the game, and therefore better understands the game, can make more sophisticated decisions than the other player.[19]

However, interestingly, compared with the case in which both players have a very high cognitive threshold, $k^1 = k^2 = 5$, payoffs for *both players* are lower when one player has a lower cognitive threshold. As described above, the player with a low cognitive threshold, player 1, learns almost all the outcomes whereas the player with a high threshold, player 2, learns outcome 1. As player 1 takes action 2, the payoff that player 2 receives from using action 1 decreases. Consequently, player 2 also learns to take action 2 to get a

[18] As noted above, for each set of parameter values, we take the average of the results generated by 100 simulation runs, while giving varying random seeds for each run.

[19] This result is in line with Josephson (2008) who has analyzed evolutionary stability of the class of action learning models that can be represented by an EWA learning model (Camerer and Ho, 1999). He found that learning rules that make little use of foregone (hypothetical) payoffs are not evolutionarily stable, i.e., they can be invaded by other learning rules. In the context of our analysis, foregone payoffs enter in the learning for those players who know the game. Consequently, players who are faster at learning about the game have an advantage in competition against slow learners.

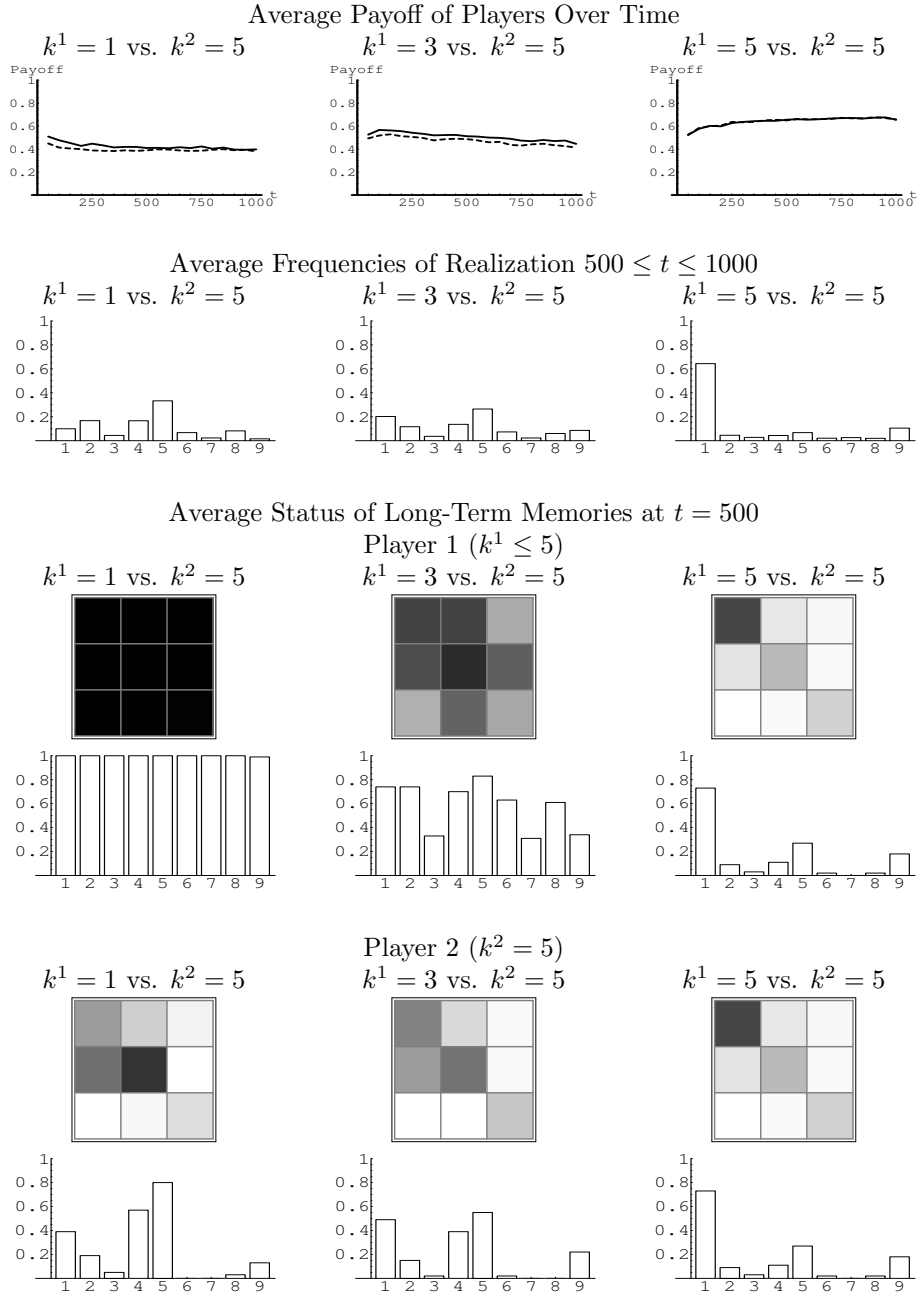

Figure 4: Average payoff of the row-player over time (top), average frequencies of realized outcomes (middle), and average status of long-term memories (bottom) for three values of k^1. $m^1 = m^2 = 5$, $k^2 = 5$, $a = 0.25$, $\lambda = 5.0$. In the top figure, the average payoff of player 1 (low cognitive threshold) is the solid line, whereas that of player 2 (high cognitive threshold) is the dashed line. For the average status in the long-term memories, the darker gray corresponds to the higher likelihood that the outcome is retained as a long-term memory. The same information is also represented by the height of bars.

higher payoff. Once both the players learn the Nash equilibrium outcome, they do not deviate from it. However, because the Nash equilibrium outcome is Pareto dominated, the payoffs of both players are lower.

Note also that when one of the players has a low k, the player with a high k also learns the Nash equilibrium outcome much more often (see the bottom panel of the figure). In addition, when a is smaller, the payoff difference between the player with a low k and the player with a high k is larger. See the appendix for more discussion.

4. Account for the Simulation Results

Why can a partial understanding of the game structure benefit players in the focal game, and why do behaviors of players remain such that their understanding of the game remains partial? In this section, we provide an explanation through a highly simplified analysis of the case where players have an identical short-term memory length and cognition threshold.

In our simulation, the model always seems to reach a state in which probabilities with which players choose their actions do not vary much over time (at least, if we take averages over several realizations). Here we restrict our analysis to such a situation. Because the game in this paper is symmetric and the players have identical characteristics (i.e., m^i, k^i, and λ^i are all the same across players), we specifically examine a symmetric case. More rigorous and exact analyses are necessary to understand fully the behavior of the model, but the simplified analysis presented below can explain the main result in our simulation analysis, namely, why partial understanding of a game can benefit players and why behaviors of players remain such that their understandings of the game remain partial.

We start with consideration of learning based only on the realized payoffs (RL model). The expected level of attraction, A_s^i, for action s in this case is

$$A_s^i = p_s^i \left(p_1^j \pi(a_s^i, a_1^j) + p_2^j \pi(a_s^i, a_2^j) + p_3^j \pi(a_s^i, a_3^j) \right), \tag{5}$$

where p_s^i and p_s^j are player i's and j's probabilities for choosing action s, respectively, as in eq. (2).

Equation (5) can be obtained is because, in RL model, an action will not receive the stimulus unless it is actually chosen. When it is chosen, the expected value of the stimulus that the action receives depends on the probabilities with which the opponent is choosing the actions. Knowing that $p_s^j = \frac{\exp(\lambda A_s^j)}{\sum_{k=1}^3 \exp(\lambda A_k^j)}$ and $A_s^i = A_s^j (\equiv A_s^{RL})$ for $s \in \{1, 2, 3\}$ in the symmetric situation, the expected levels of attraction for the RL model (A_s^{RL}) become

$$A_1^{RL} = p_1^{RL}(p_1^{RL}(1-a) + p_3^{RL}),$$
$$A_2^{RL} = p_2^{RL}(p_1^{RL} + p_2^{RL}a + p_3^{RL}a),$$
$$A_3^{RL} = p_3^{RL}(p_1^{RL} + p_3^{RL}(1-a)),$$

where $p_s^{RL} = \frac{\exp(\lambda A_s^{RL})}{\sum_{k=1}^3 \exp(\lambda A_k^{RL})}$ for each action s in the game. Here we have ignored adding b to the payoffs for clarity of exposition. Solving these equations for A_1^{RL}, A_2^{RL}, and A_3^{RL} gives us the expected levels of attractions for each action, and the expected probabilities that players choose each action follows immediately.[20]

In our model of game learning, players' understanding of the game plays a role in determining the expected level of attraction. For example, if the payoff associated with (s_1^i, s_1^j) is a unique long-term memory in the players' minds, then the expected attraction of our model (A_s^{GL}) become

$$A_1^{GL} = p_1^{GL}(1-a) + p_1^{GL}p_3^{GL},$$
$$A_2^{GL} = p_2^{GL}(p_1^{GL} + p_2^{GL}a + p_3^{GL}a),$$
$$A_3^{GL} = p_3^{GL}(p_1^{GL} + p_3^{GL}(1-a)).$$

These expressions can be obtained because action 1 will elicit a stimulus not only when

[20] It is possible that multiple solutions exist.

outcomes (s_1^i, s_1^j), (s_1^i, s_2^j), and (s_1^i, s_3^j) are realized but also every time the opponent chooses action 1. One can expect from these equations that when players only learn of the payoffs associated with (s_1^i, s_1^j), action 1 will have a higher expected attraction than the RL model and will be chosen with a higher probability by players. Consequently, (s_1^i, s_1^j) will be observed much more frequently in our model than in the case of the RL model. Furthermore, because players choose other actions with a low probability, other outcomes are not realized frequently enough. Therefore, their understanding of the game remains partial.

On the other hand, if players learn the entire game quickly, as in the $k = 1$ case, the expected attractions of our model immediately become equivalent to those of learning based on both realized and forgone payoffs (FP model). The expected attraction for action s in the FP model is given as

$$A_s^i = p_1^j \pi(a_s^i, a_1^j) + p_2^j \pi(a_s^i, a_2^j) + p_3^j \pi(a_s^i, a_3^j). \tag{6}$$

Equation (6) can be derived because all the actions will always receive stimulus, irrespective of whether they have been chosen or not. Therefore, the expected levels of attractions in the FP model (A_s^{FP}) become

$$\begin{aligned} A_1^{FP} &= p_1^{FP}(1-a) + p_3^{FP}, \\ A_2^{FP} &= p_1^{FP} + p_2^{FP} a + p_3^{FP} a, \\ A_3^{FP} &= p_1^{FP} + p_3^{FP}(1-a), \end{aligned}$$

for the game when we ignore the b added to the payoffs. These expressions give the largest weight to choosing action 2 for both of the players. Therefore, compared with the RL model, the FP model results in players obtaining lower payoffs.

5. Summary and Conclusion

This paper presented a model of learning about a game. Players initially have little knowledge about the game they play. They gain experience through playing the game repeatedly. Based on their experience, they not only learn which action will bring about a higher payoff but also form their view about the game they are playing. We show that, in the 3×3 game we have considered, which embeds both a prisoner's dilemma and a coordination game, players might benefit from having a very limited understanding of the game when all the players have such a limited understanding. Their payoffs can be higher than in the cases where players have full or no understanding of the game. It is noteworthy that personal views and behaviors of players coevolve in our model. Players enjoy a high payoff. Therefore, their understanding of the game remains partial and vice versa. This result suggests that players might live happily without fully understanding highly complex strategic environments.

When one player has a much better understanding of the game than the other player, the one with the better understanding can enjoy a higher payoff than the player with less understanding. However, their payoffs—even the payoffs of the player who better understands the game—can be lower than in the case in which all the players have a limited understanding. The behavior of the player who better understands the game can lead the other player to respond in a way that lowers their payoffs. This, combined with the results above, suggests that a benefit of ignorance might exist, but it exists only when everyone is ignorant.

Our paper has been motivated the recent development of a new theory, *inductive game theory*, by Kaneko and Kline (2007, 2008).[21] Inductive game theory investigates experiential foundations of beliefs/knowledge with players ' memories. Akiyama et al. (2008) represents two types of memories, short-term and long-term memories, based on informal theory in Kaneko and Kline (2008), and show boundary of learning within practical finite time. On the other hand, this paper concentrates on explicit analysis of

[21] See §9.3 in Kaneko and Kline (2008) for details.

learning dynamics when giving a probabilistic behavior model with the same types of memories in Akiyama et al. (2008). In contrast with Akiyama et al. (2008), this paper considers the relationship between players ' cognitive abilities and average payoffs in a two-person game. As a result, we found that the interaction between players with different cognitive abilities has an influence of each player 's average payoff, as described above.

As described herein, we have considered a pair of players playing the game repeatedly. However, one can easily extend the framework presented here to the case where there are many players to be matched with a few others. In such a case, it is possible to consider various matching protocols: for example, players might be situated in a network and interact only locally. The results here suggest that it is possible that players form several different "local views" of the same objective game. What will happen when occasional random matching exists among those with different views? Are there views that can spread much more easily than others? These are all interesting questions to investigate, but we will leave them for future research.

It is also interesting to conduct laboratory experiments and examine how subjects learn in the situation considered in this paper; namely, subjects are initially only informed about the set of actions available to themselves, but they observe actions chosen by all the relevant players and the payoff received after each interaction. Do subjects behave in the way that the model predicts? We leave these questions also for future research.

Acknowledgements

We thank Mamoru Kaneko, Jeff Kline, Daisuke Oyama, and seminar participants at GREQAM for comments and suggestions. This research is partially supported by a Japan Society for the Promotion of Sciences Grant-in-Aid for Scientific Research (S), No. 17103002, and by Ministry of Education, Culture, Sports, Science and Technology Grants-in-Aid for Young Scientists (B), No. 19730137 and No. 19730164.

A. Dependency of the Results on Parameter Values

In the main text, we presented results under $m^1 = m^2 = 5$, $a = 0.25$, $\lambda = 5.0$. Here we show what happens to the results if we change the parameter values.

Figures 5 and 6 show results for the case in which $a = 0.05$ and $a = 0.45$, respectively, in the same format as Fig. 2. As in the case discussed in the main text, when $k = 1$, GL players quickly learn all the outcomes, and their behaviors converge to that of FP players. When $a = 0.45$, there is not much difference in behavior among the GL, FP, and RL players. It is also the case in which among GL players, the differences in cognition threshold k do not affect their behavior markedly. In all the models, players learn to play the Nash equilibrium. Then they learn the payoff associated with the Nash equilibrium outcome almost all the time.

In these two figures, the GL players do not receive higher payoffs than FP or RL players, contrary to the description in the main text and Fig. 2. What is the range of a over which our main result holds? How about the range of λ? Figure 7 presents the average payoff of players for various values of a holding λ constant at $\lambda = 5.0$ (top) as well as various values of λ while holding a constant at $a = 0.25$. The GL players receive higher payoffs than FP and RL players over quite a large parameter space: in particular, $0.1 \leq a \leq 0.35$ when $\lambda = 5.0$ and $3.0 \leq \lambda \leq 6.0$ for $a = 0.25$. In fact, we have experimented with other values of m, and have obtained similar results: as k becomes closer to m, the GL players receive high payoffs while their understanding of the game remains very limited, although the specific values of a and λ for which such a result holds depend on m.

When players have the same short-term memory length but different cognition thresholds, the player with a low cognition threshold (the one who learns the game quickly) receives a higher payoff than the one with a high cognition threshold. The difference between the payoffs received by the two players is larger when the difference between the two thresholds is large and also when a is low.

Figure 8 presents results for simulation runs when $m^1 = m^2 = 5$, $\lambda = 5.0$ and

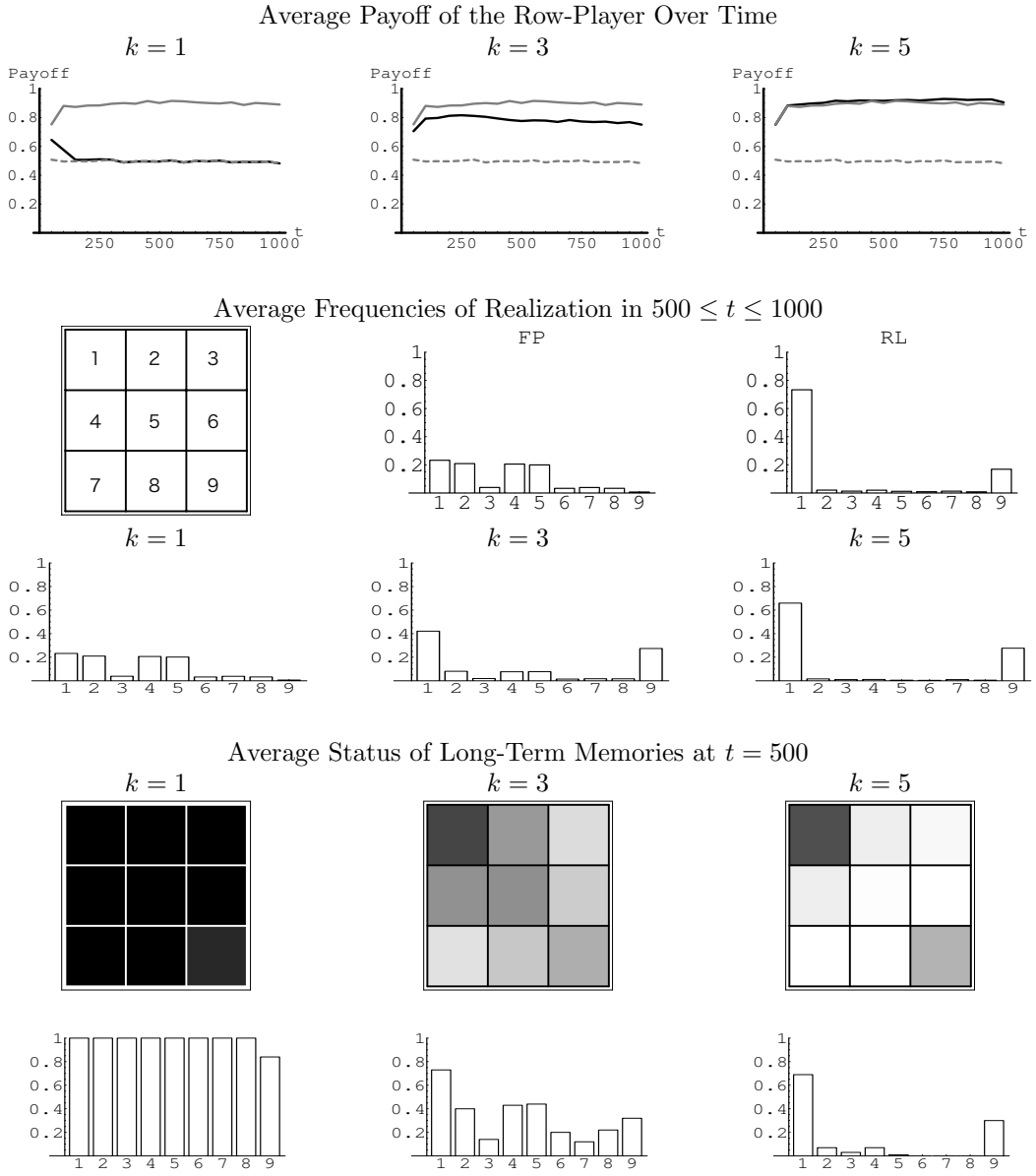

Figure 5: Average payoff of the row-player over time (top), average frequencies of realized outcomes (middle), and average status of long-term memories (bottom) for various k. $m = 5$, $a = 0.05$, $\lambda = 5.0$. For the average payoff, the result of our model is in solid black, the solid gray represents the RL model, and the dashed gray represents the FP model. For the average status in the long-term memory, the darker gray corresponds to the higher likelihood that the outcome is retained as a long-term memory, which is also represented by the height of bars.

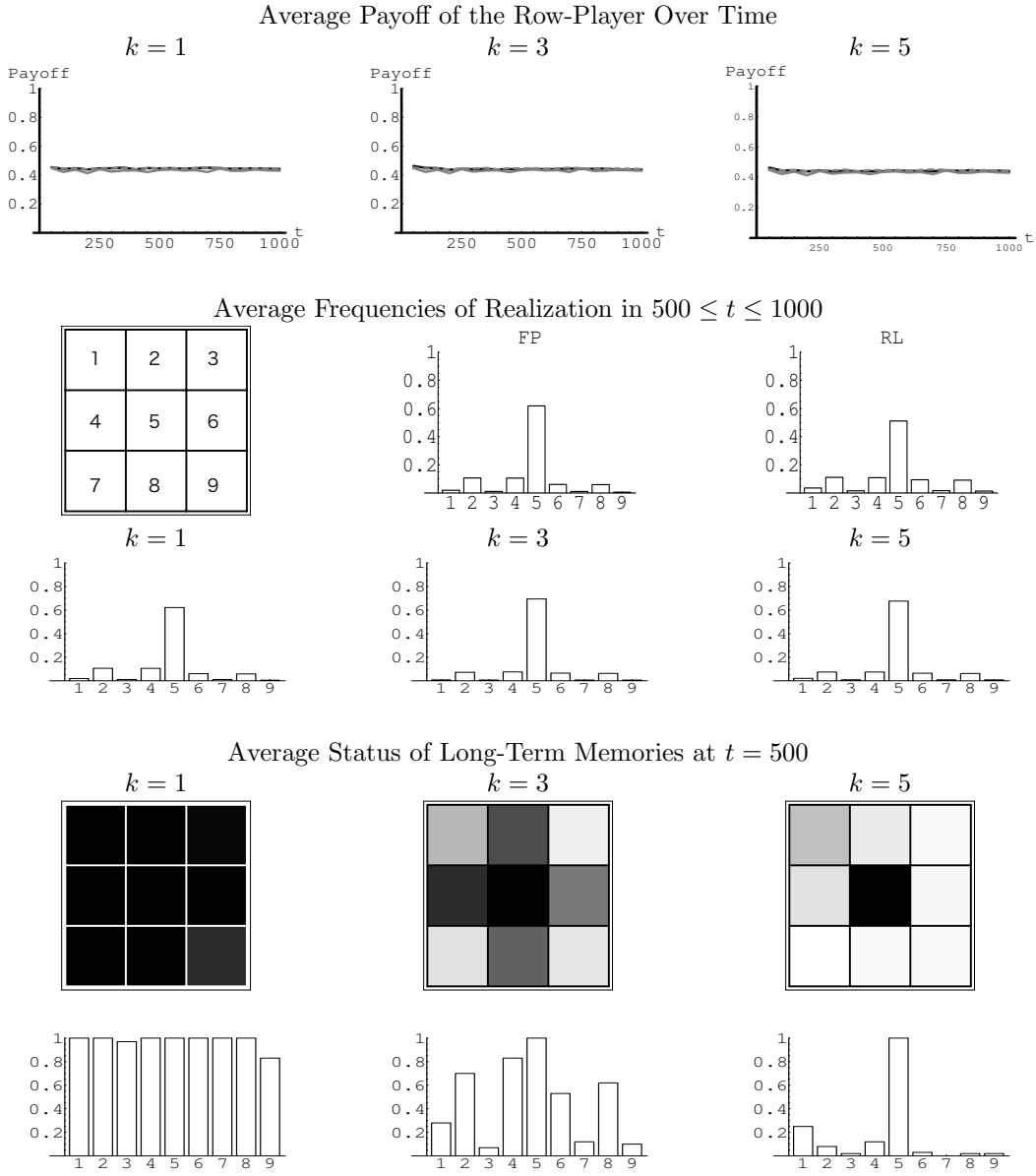

Figure 6: Average payoff of the row-player over time (top), average frequencies of realized outcomes (middle), and average status of long-term memories (bottom) for various k. $m = 5$, $a = 0.45$, and $\lambda = 5.0$. For the average payoff, the result of our model is shown in solid black. Solid gray represents the RL model; the dashed gray represents the FP model. For the average status of long-term memories, the darker gray corresponds to the higher likelihood that the outcome is retained as a long-term memory. The same information is also represented by the height of bars.

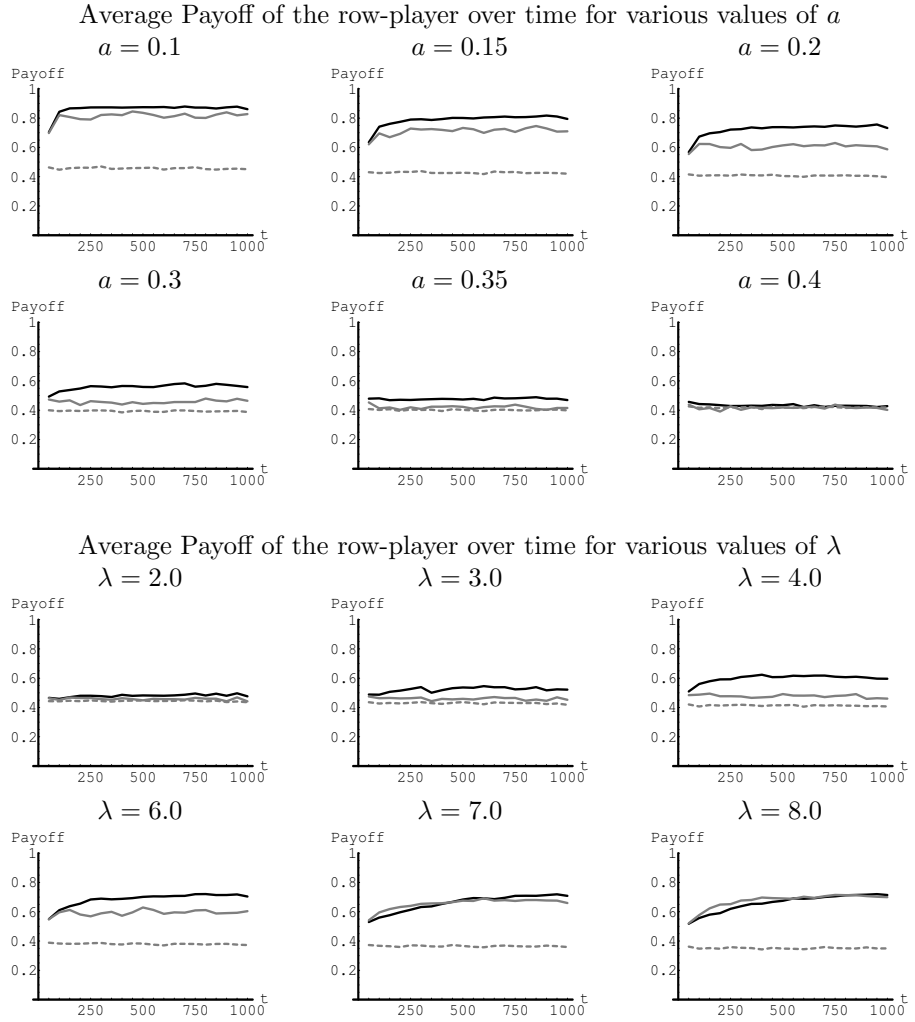

Figure 7: Average payoff of the row-player over time for various a (top) and for various values of λ (bottom). $m = 5$, $k = 5$, $\lambda = 5.0$ (top) and $a = 0.25$ (bottom). For the average payoff, the result of our model is presented in solid black. The solid gray represents RL model. The dashed gray represents the FP model.

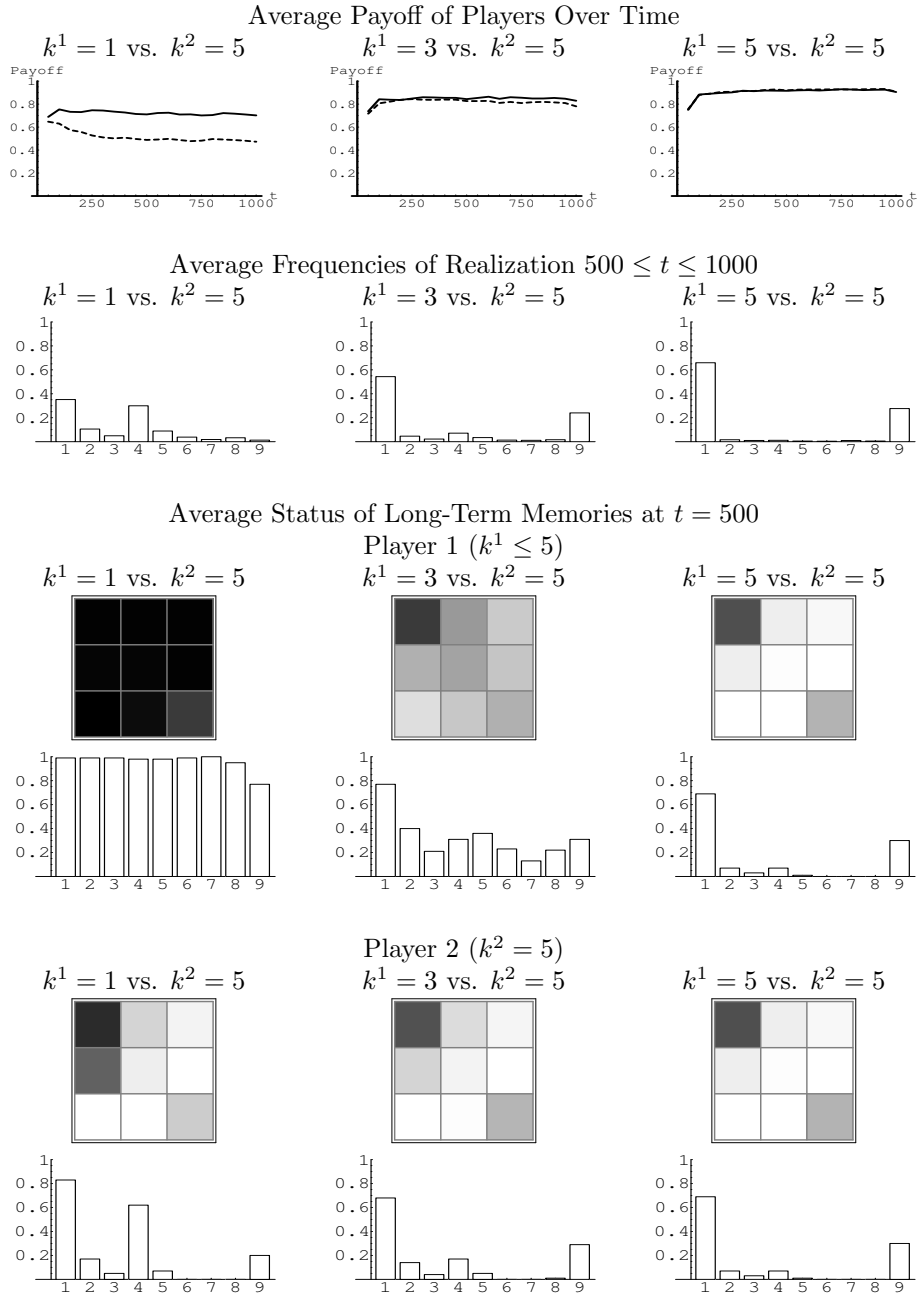

Figure 8: Average payoff of the row-player over time (top), average frequencies of realized outcomes (middle), and average status of long-term memories (bottom) for three values of k^1. $m^1 = m^2 = 5$, $k^2 = 5$, $a = 0.05$, $\lambda = 5.0$. In the top panel, the average payoff of player 1 (low cognitive threshold) is shown as a solid line, whereas that of player 2 (high cognitive threshold) is shown as the dashed line. For the average status of the long-term memories, the darker gray corresponds to the higher likelihood that the outcome is recorded as a long-term memory. The same information is also represented by the height of bars.

$a = 0.05$. As might be readily apparent, the lower k^1 is, the greater the payoff difference between the two players. Furthermore, compared with what was shown in Fig. 4, the payoff differences between the two players are larger here.

B. Longer Run Results

The results presented in the main text were based on the first 1000 periods of simulation because our interest is not necessarily in the convergent states but in the construction of personal view by the players with limited abilities. It is noteworthy that because of the construction of the model, in particular the probabilistic action choices, the longer the simulation is, the better players learn about the game. In fact, if the game was going to be repeated infinitely many times, then GL players would eventually learn all the payoffs, and their behaviors would become equivalent to those of FP players.

In a finite time horizon, however, how much GL players can learn about the game depends on their short-term memory length, m, and cognition threshold, k, as well as other parameters of the model. To illustrate this point, the average payoffs obtained by GL players (black), RL players (solid gray), and FP players (dashed gray) over 50 000 periods are portrayed in Fig. 9. In the figure, results for GL players of two types are described, $m = k = 5$ (left) and $m = k = 10$ (right).[22] Here two players with the same short-term memory length and cognition thresholds are matched. Other parameters are the same as in Fig. 2. As shown in the figure, a declining trend exists, in the long run, for the average payoff of GL players in case of $m = k = 5$. On the contrary, we cannot observe such an explicit trend in the case of $m = k = 10$. The difference between these two results arises from the ease with which players learn about the payoffs. The higher the cognition threshold, the more difficult it is for players to learn. Consequently, it takes much longer for their behaviors (therefore, payoffs) to converge to those of FP players.

[22] Just as in all the simulations presented in this paper, FP players and RL players have the same short-term memory length as GL players.

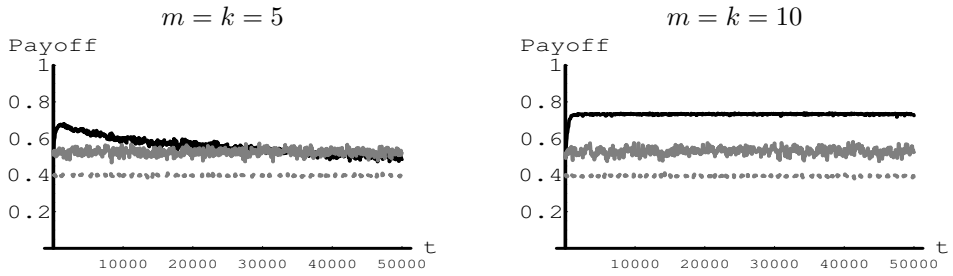

Figure 9: Average payoff of the row-player over time. Both players having $m = k = 5$ (left) and $m = k = 10$ (right). $a = 0.25$, $\lambda = 5.0$. The GL model is shown in solid black. The solid gray represents the RL model. The dashed gray represents the FP model. The payoffs are averaged over 100 simulation runs.

C. Results from 2×2 games

In the main text, we presented results from 3×3 games that embed both prisoner's dilemma and coordination games. We have specifically examined such games because we were interested in studying what kind of personal views emerge from our model. Our model, however, is applicable to simpler 2×2 games as well. In this appendix, we show the results from applying GL model as well as RL and FP models to the prisoner's dilemma game and coordination game separately. The results discussed here show that in both the prisoner's dilemma game and coordination game with risk-payoff tradeoff, we obtain similar outcomes in terms of the average payoffs in the first 1000 periods. That is, GL players with a high cognition threshold might benefit from their limited understanding of the game, and might obtain higher payoffs than those players with no understanding of the game (RL players) or full understanding of the game (FP players).

In this section, we assume that two players with the same short-term memory length, $m = 5$, and cognition threshold, k, are matched to play the game. We consider three cognition thresholds $k \in \{1, 3, 5\}$. We have also set the sensitivity of action choices to the attraction, $\lambda = 5.0$, as in most of the simulations reported in the main text.

Figure 10 presents results of our simulation for a prisoner's dilemma game with $a =$

0.25. The average payoffs demonstrate similar results to those obtained in the 3×3 game. Namely, GL players with a high cognition threshold, $k = 5$, obtain higher average payoffs than both RL and FP players. In addition, as before, RL players obtain higher payoff than FP players. The GL players with lower cognition thresholds learn about the game relatively quickly and their behaviors converge to those of FP players. The average status of long-term memories shows somewhat different results from the 3×3 games. (Four outcomes that correspond to prisoner's dilemma are 1, 2, 4, 5 in Fig. 2.) Namely, the frequency for outcome 1 (efficient outcome) being retained as a long-term memory is lower, and that for outcome 4 (outcome 5 in the case of 3×3 game) is higher than the case considered in main text.

Figure 11 presents results for a coordination game with $a = 0.35$. We have chosen $a = 0.35$ for this game instead of $a = 0.25$ because the risk–payoff tradeoff takes place only for $1/3 < a < 0.5$.[23] For this game, the average payoffs demonstrate similar results as in the 3×3 game. Namely, GL players with high cognition threshold, $k = 5$, obtain higher average payoffs than both RL and FP players. The RL model results in a payoff-dominant equilibrium (lower right outcome, outcome 4) more frequently than the risk-dominant equilibrium (upper left outcome, outcome 1), whereas it is opposite in the case of FP model. The GL model, when the cognition threshold is high, results in the payoff dominant equilibrium more frequently than the RL model.

[23]For $a = 0.25$, the lower right outcome is realized most frequently in all the models described in this paper.

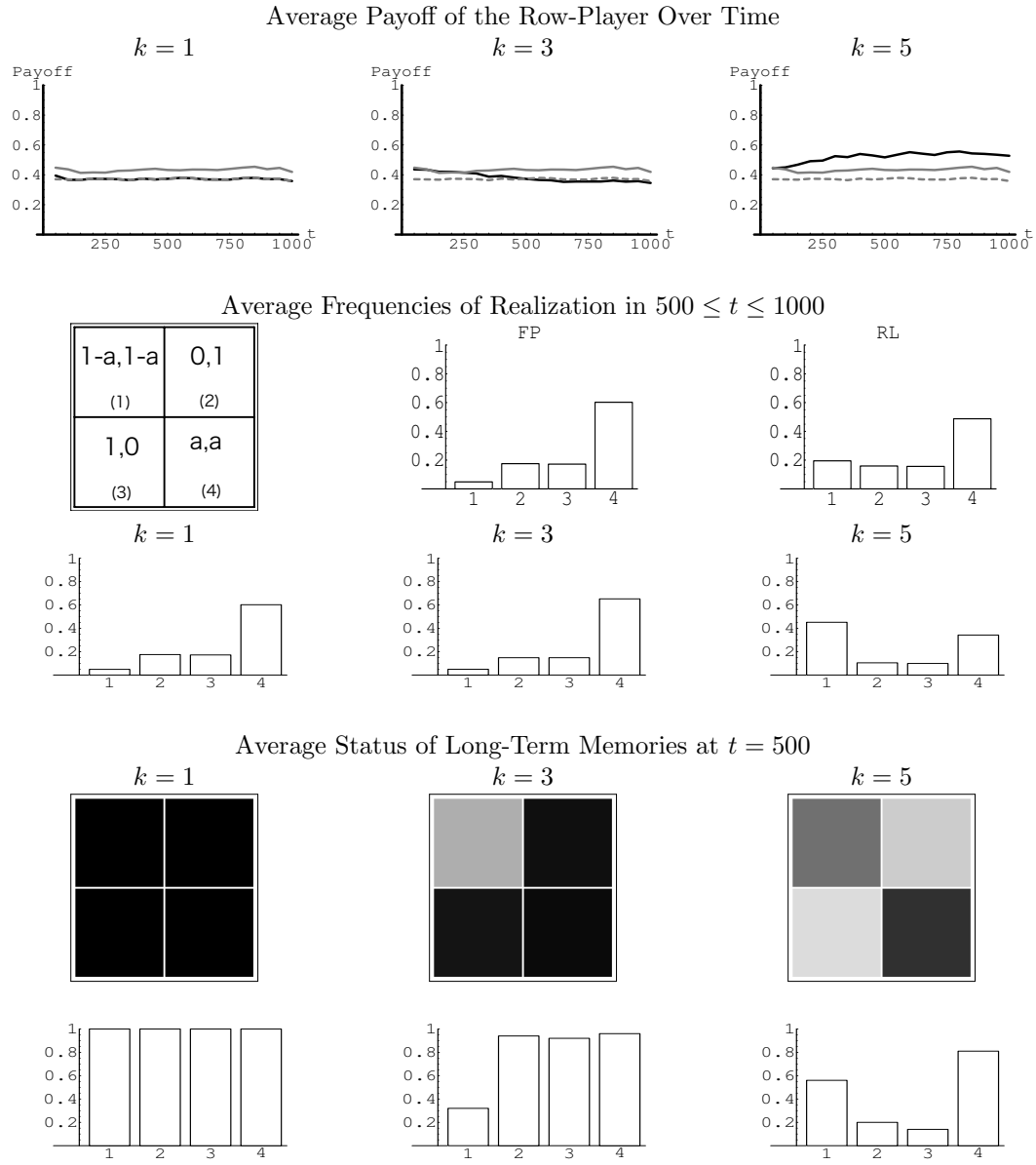

Figure 10: A result from the prisoner's dilemma game $a = 0.25$. The average payoff of the row-player over time (top), average frequencies of realized outcomes (middle), and average status of long-term memories (bottom) for various k. $m = 5$, $\lambda = 5.0$. For the average payoff, the result of our model is presented in solid black. The solid gray represents the RL model. The dashed gray represents the FP model. For the average status of long-term memories, the darker gray corresponds to the higher likelihood that the outcome is retained as a long-term memory, which is also represented by the height of bars.

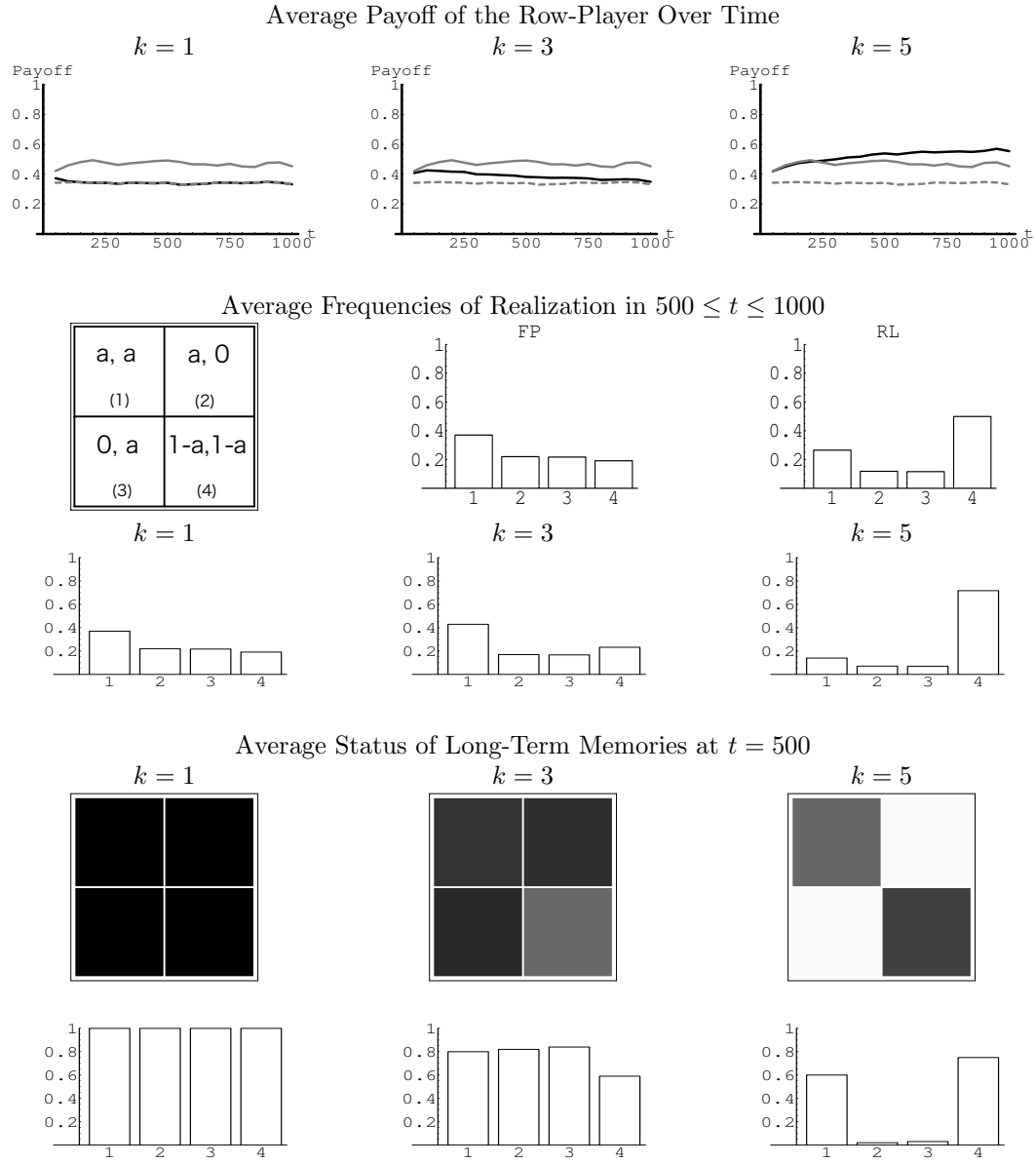

Figure 11: A result from coordination game $a = 0.35$. Average payoff of the row-player over time (top), average frequencies of realized outcomes (middle), and average status of long-term memories (bottom) for various k. $m = 5$, $\lambda = 5.0$. For the average payoff, the result of our model is in solid black, the solid gray represents the RL model, and the dashed gray represents the FP model. For the average status of long-term memories, the darker gray corresponds to the higher likelihood that the outcome is retained as a long-term memory, which is also represented by the height of bars.

References

Akiyama, E., Ishikawa, R., Kaneko, M., Kline, J. J., 2008. A simulation study of learning a structure: Mike's bike commuting. to appear in Economic Theory, available at `http://www.sk.tsukuba.ac.jp/SSM/libraries/pdf1176/1190.pdf`.

Arifovic, J., McKelvey, R. D., Pevnitskaya, S., 2006. An initial implementation of the turing tournament to learning in two person games. Games and Economic Behavior 57, 93–122.

Atkinson, R., Shiffrin, R., 1968. Human memory: A proposed system and its control processes. In: Spence, K., Spence, J. (Eds.), The psychology of learning and motivation: Advances in research and theory. Vol. 2. Academic Press., pp. 89–195.

Brewer, W. F., 1986. What is autobiographical memory? In: Rubin, D. (Ed.), Autobiographical memory. Cambridge University Press, pp. 25–49.

Brock, W. A., Hommes, C. H., 1997. A rational route to randomness. Econometrica 65 (5), 1059–1095.

Brock, W. A., Hommes, C. H., 1998. Heterogeneous beliefs and routes to chaos in a simple asset pricing model. Journal of Economic Dynamics and Control 22, 1235–1274.

Camerer, C., Ho, T.-H., 1999. Experience-weighted attraction learning in normal form games. Econometrica 67, 827–874.

Camerer, C. F., 2003. Behavioral Game Theory: Experiments in Strategic Interaction. Russell Sage Foundation, New York, NY.

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

Crawford, V. P., 1995. Adaptive dynamics in coordination games. Econometrica 63, 103–143.

Erev, I., Roth, A. E., 1998. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. American Economic Review 88, 848–881.

Fudenberg, D., Levine, D. K., 1998. The Theory of Learning in Games. MIT Press, Cambridge, MA.

Hanaki, N., Sethi, R., Erev, I., Peterhansl, A., 2005. Learning strategy. Journal of Economic Behavior and Organization 56, 523–542.

James, W., 1890. The principles of psychology. Henry Holt and Co.

Josephson, J., 2008. A numerical analysis of the evolutionary stability of learning rules. Journal of Economic Dynamics and Control, 1569–1599.

Kaneko, M., Kline, J. J., 2007. Information protocols, and extensive games in inductive game theory. Journal of Mathematics, Game Theory, and Algebra 17 (5/6).

Kaneko, M., Kline, J. J., 2008. Inductive game theory: A basic scenario. Journal of Mathematical Economics 44, 1332–1363.

Linton, H. A., 1982. Transformations of memory in everyday life. In: Neisser, U. (Ed.), Memory observed: Remembering in natural contexts. Freeman.

McKelvey, R. D., Palfrey, T. R., 1995. Quantal response equilibria for normal form games. Games and Economic Behavior 10, 6–38.

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

Oechssler, J., Schipper, B., 2003. Can you guess the game you are playing? Games and Economic Behavior 43, 137–152.

Tulving, E., 1972. Episodic and semantic memory. In: Tulving, E., Donaldson, W. (Eds.), Organization of memory. Academic Press, New York, pp. 381–403.

Wagenaar, W. A., 1986. My memory: A study of autobiographical memory over six years. Cognitive Psychology 18, 225–252.

Waltman, L., Kaymak, U., 2008. Q-learning agents in a cournot oligopoly model. Journal of Economic Dynamics and Control 32, 3275–3293.