

Repeated games potentially include memories of agents, i.e., a possibility that agents determine their actions depending on past actions they chose (see Fig. 1 for the illustration). Such memories can expand the choice of strategies and thus lead to the agents handling their gameplay better; for example, by reading how the other player chooses its action [19]. Indeed, agents with memories can use tit-for-tat [20] and win-stay-lose-shift [21] strategies in prisoner’s dilemma games, and these strategies achieve cooperation as Nash equilibrium, explaining human behaviors. Furthermore, how a region of the Nash equilibrium is extended by multi-memory strategies is enthusiastically studied as folk theorem [22]. In practice, Q-learning is frequently implemented in multi-memory games [23, 24, 25]. Several studies [26, 27] partly discuss the relation between the replicator dynamics and the gradient ascent but consider only prisoner’s dilemma games. In conclusion, this relation is still unclear in games with general numbers of memories and actions. Furthermore, the convergence of dynamics in such multi-memory games has been unexplored.

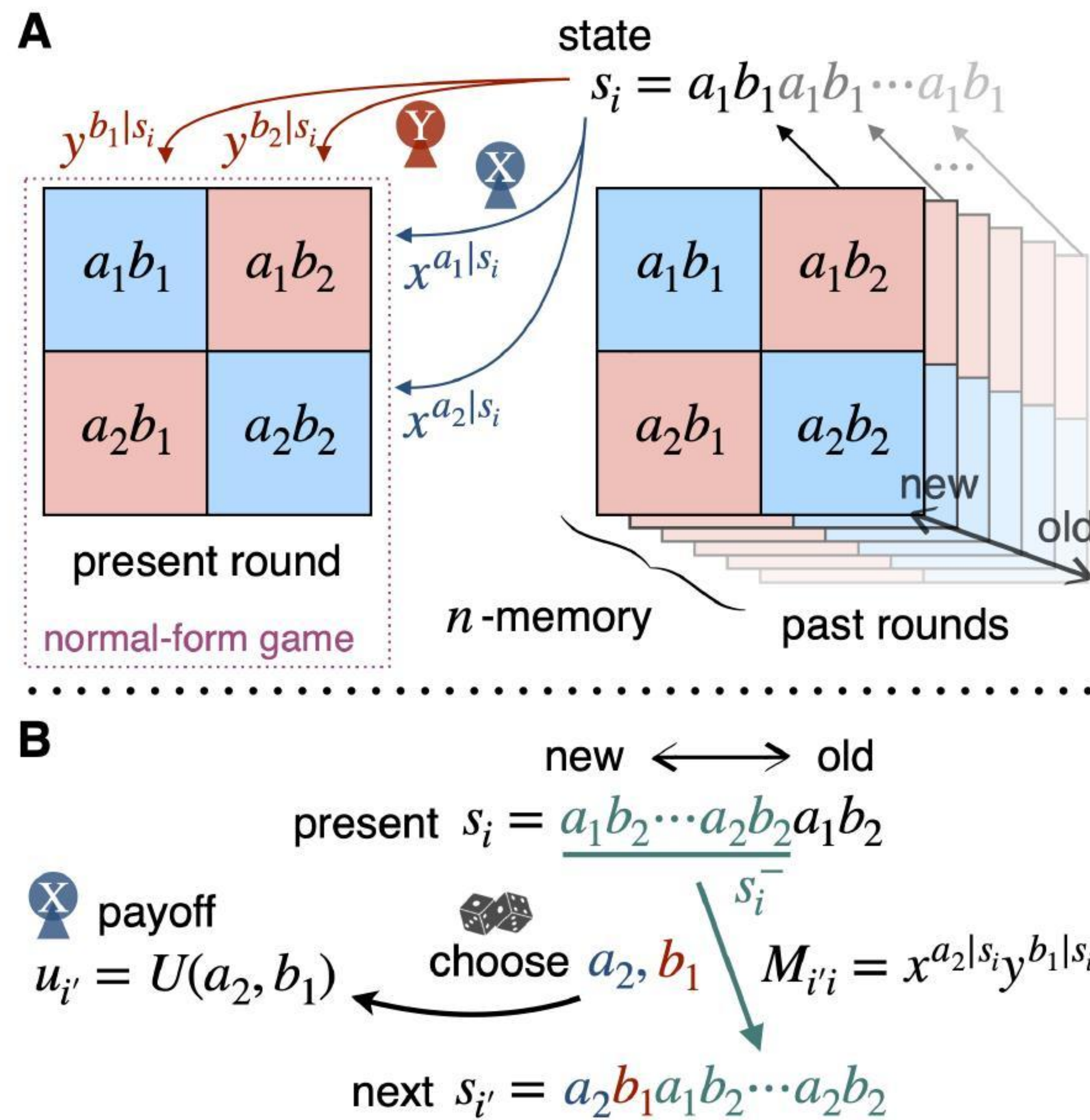


Figure 1: **A.** Illustration of a multi-memory repeated game. Focusing on the area surrounded by the purple dots, a normal-form game is illustrated. Player X (resp. Y) chooses its action a_1 or a_2 in the row (resp. b_1 or b_2) in the column. Then, each of them receives its payoff depending on their actions. The panel shows a penny-matching game, where blue (resp. red) panels show that X (resp. Y) gains a payoff of 1 and the other loses it. Looking at the whole, each player memorizes their actions of the past n rounds. This memorized state is described as s_i given by $2n$ -length bits of actions. **B.** Illustration for the detailed single round of repeated games, where present state s_i transitions to next state $s_{i'}$. In this transition, the oldest 2 bits are lost, and the other bits s_i^- , colored in green, are maintained. X’s and Y’s choices (a_2 (blue) and b_1 (red) in this figure) are appended as the newest 2 bits in $s_{i'}$. This transition occurs with the probability of $M_{i'i}$. Finally, X gains a payoff of $u_{i'}$ in the state transition.

This study provides a basic analysis of the multi-memory repeated game. First, we extend the two learning algorithms, i.e., replicator dynamics and gradient ascent, for multi-memory games. Then, we name them multi-memory replicator dynamics (MMRD) and gradient ascent (MMGA). As well as shown in the zero-memory games, the equivalence between MMRD and MMGA is proved in Theorems 1-3. Next, we tackle the convergence problem of such algorithms from both viewpoints of theory and experiment. Theorem 4 shows that Nash equilibrium uniquely exists in multi-memory zero-sum games as well as zero-memory ones. This theorem is nontrivial if taking into account the fact that diversification of strategies can expand the region of Nash equilibrium in general games. Then, while utilizing these theorems, we see how multi-memory learning complicates the dynamics, leading to divergence from the Nash equilibrium with sensitivity to its initial condition like chaos.