

Assumption 1 (One-memory two-action zero-sum game). We assume a two-action (i.e., $\mathcal{A} = \{a_1, a_2\}$ and $\mathcal{B} = \{b_1, b_2\}$), one-memory (i.e., $\mathbf{s} = (a_1b_1, a_1b_2, a_2b_1, a_2b_2)$), and zero-sum game (i.e., $\mathbf{v} = -\mathbf{u}$). In particular, we discuss zero-sum games where both u_1 and u_4 are smaller or larger than both u_2 and u_3 .

Under Assumption 1, we exclude uninteresting zero-sum payoff matrices that the Nash equilibrium exists as a set of pure strategies because the learning dynamics trivially converge to such pure strategies. The condition that both u_1 and u_4 are smaller or larger than both u_2 and u_3 is necessary and sufficient for the existence of no dominant pure strategy.

In the rest of this paper, we use a vector notation for strategies of X and Y; $\mathbf{x} := \{x_i\}_{i=1,\dots,4}$ and $\mathbf{y} := \{y_i\}_{i=1,\dots,4}$ as $x_i := x^{a_1|s_i}$ and $y_i := y^{b_1|s_i}$. Indeed, $x^{a_2|s_i} = 1 - x_i$ and $y^{b_2|s_i} = 1 - y_i$ hold.

Theorem 4 (Uniqueness of Nash equilibrium). Under Assumption 1, the unique Nash equilibrium of this game is $(x_i, y_i) = (x^*, y^*)$ for all i as

$$x^* = \frac{-u_3 + u_4}{u_1 - u_2 - u_3 + u_4}, \quad y^* = \frac{-u_2 + u_4}{u_1 - u_2 - u_3 + u_4}. \quad (11)$$

Proof Sketch. Let us prove that X's strategy in Nash equilibrium is uniquely $\mathbf{x} = x^*\mathbf{1}$. First, we define u^* and v^* as X's and Y's payoffs in the Nash equilibrium in the zero-memory game. If $\mathbf{x} = x^*\mathbf{1}$, X's expected payoff is $u^{\text{st}} = u^*$, regardless of Y's strategy \mathbf{y} . Second, we consider that X uses another strategy $\mathbf{x} \neq x^*\mathbf{1}$. Then, there is Y's strategy such that $v^{\text{st}} > v^* \Leftrightarrow u^{\text{st}} < u^*$. Thus, X's minimax strategy is uniquely $\mathbf{x} = x^*\mathbf{1}$, completing the proof. See Appendix A.4 for the full proof. \square

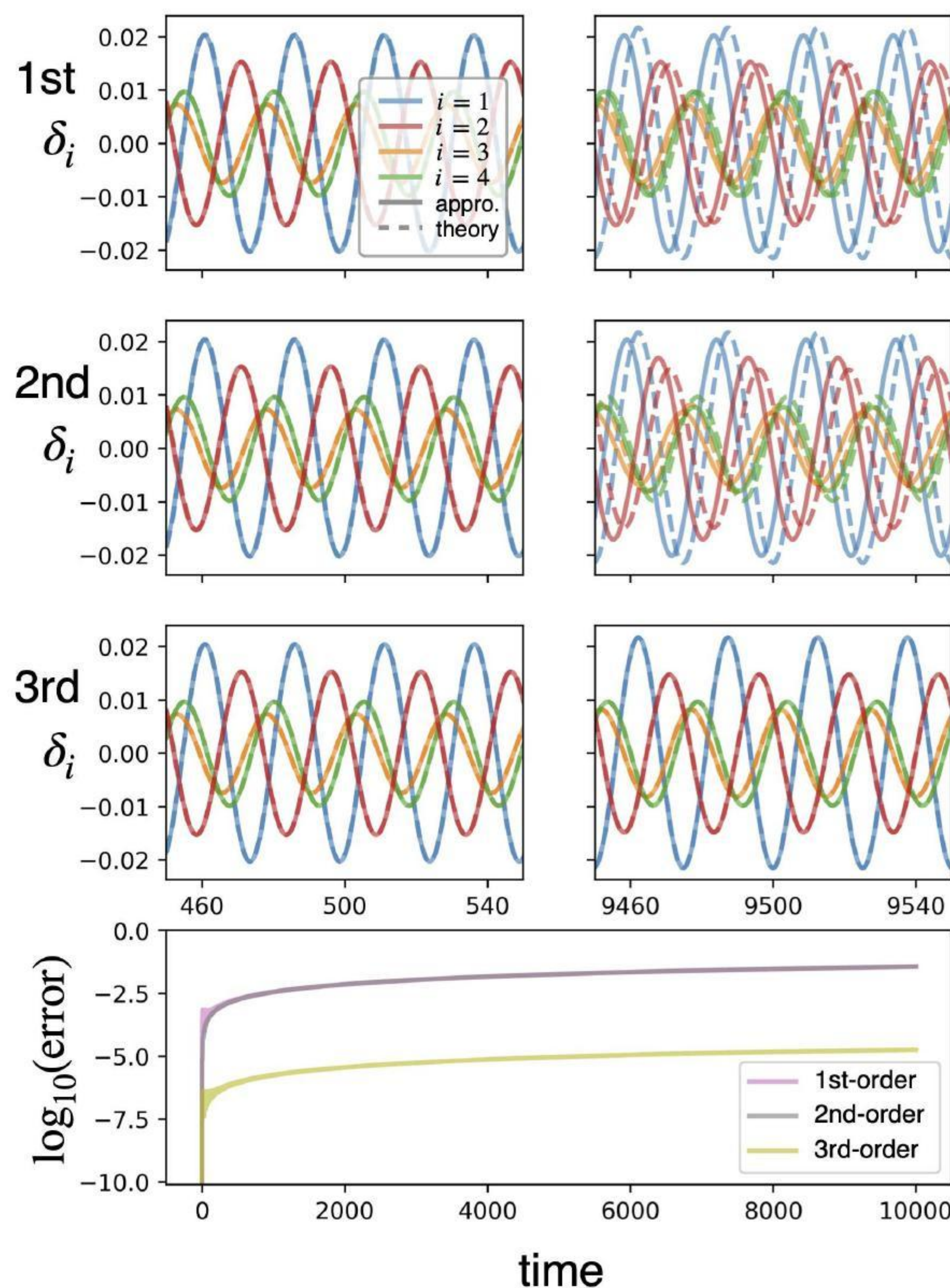


Figure 2: Multi-memory learning dynamics near the Nash equilibrium in the penny-matching game. In the upper six panels, colored lines indicate the time series of δ_i (X's strategy). The solid (resp. broken) lines are approximated (resp. experimental) trajectories of learning dynamics. From the top, the trajectories are predicted by approximations up to the first, second, and third orders. The bottom panel shows the errors between the approximated and experimental trajectories.

Regarding Theorem 4, X (Y) chooses each action in the same probability independent of the last state. Here, they do not utilize their memory. Thus, note that in this sense, the Nash equilibrium is the same as