Figure 3: Initial state sensitivity in learning dynamics in multi-memory games. In the top panels, colored lines are time series of x_i (X's strategy). The black line is the distance between the solid (sample of $\boldsymbol{x}$) and broken ($\boldsymbol{x}'$) lines. In the bottom panels, the black lines indicate the maximum eigenvalue in the learning dynamics of the solid line.
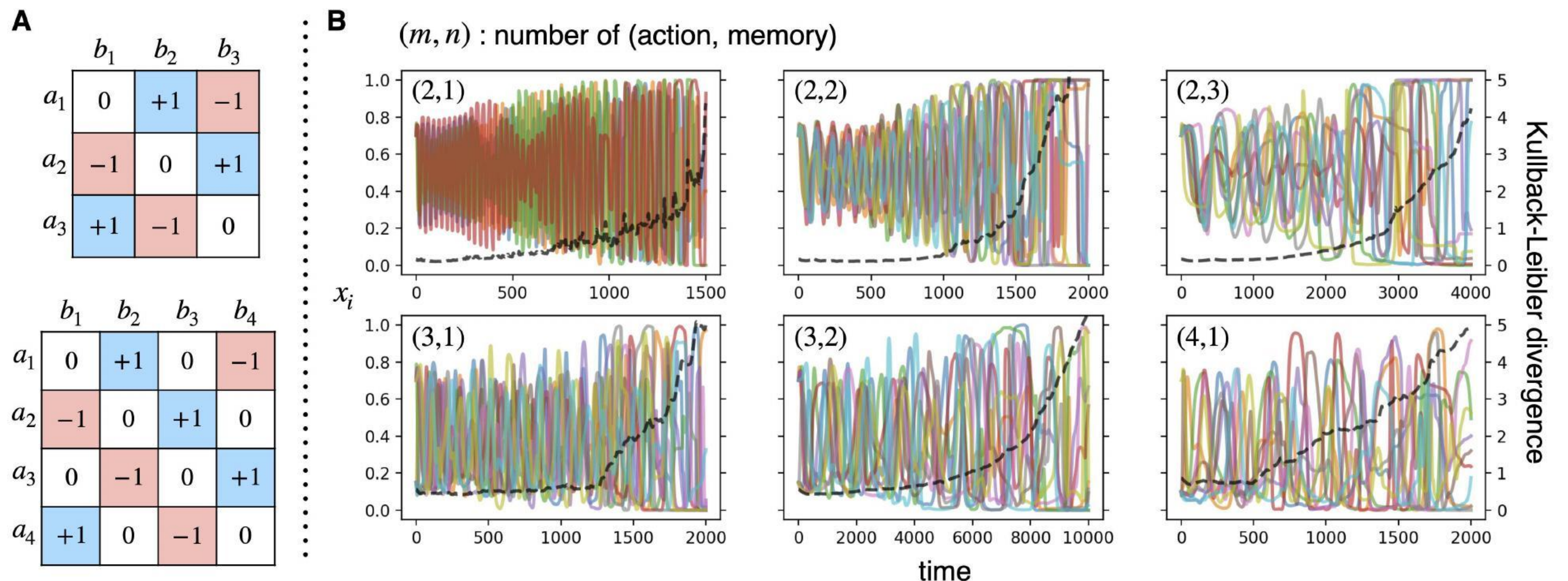

Figure 4: **A**. Payoff matrices of three-action (rock-paper-scissors) and four-action (extended rock-paper-scissors) games. **B**. In each panel, colored lines indicate time series of $x^{a|s}$ for random $a \in \mathcal{A}$ and $s \in \mathcal{S}$. The black broken line indicates the Kullback-Leibler divergence averaged over all the states $s \in \mathcal{S}$, intuitively meaning a distance from the Nash equilibrium.

that in the zero-memory version of the game. This theorem means that in zero-sum games, the region of Nash equilibrium does not expand even if players have memories. Taking into account that having multiple memories expands the region of Nash equilibrium, such as a cooperative equilibrium in prisoner's dilemma games [20], this theorem is nontrivial.

In order to discuss whether our algorithms converge to this unique Nash equilibrium under Assumption 1, we consider the neighbor of Nash equilibrium and define sufficient small deviation from the Nash equilibrium, i.e., $\boldsymbol{\delta} := \boldsymbol{x} - x^*\boldsymbol{1}$ and $\boldsymbol{\epsilon} := \boldsymbol{y} - y^*\boldsymbol{1}$. Here, we assume that these deviations have the same scale $O(\delta) := O(\delta_i) = O(\epsilon_i)$ for all i. Then, defining that the superscript (k) shows $O(\delta^k)$ terms, the dynamics are