3 Algorithm

In the following, we define multi-memory versions of two major learning algorithms, i.e., replicator dynamics and gradient ascent. Although we consider the learning of player X, that of player Y can be formulated in the same manner.

Definition 1 (expected future payoff)**.** *We define the expected future payoff from the distribution* $\boldsymbol{p}$ *as*

$$\pi(\boldsymbol{p}, \mathbf{x}, \mathbf{y}) := \sum_{t=0}^{\infty} \boldsymbol{M}^t (\boldsymbol{p} - \boldsymbol{p}^{\mathrm{st}}) \cdot \boldsymbol{u}, \tag{5}$$

which is the total payoff player X obtains from the present round to the future.

In this definition, the stationary payoff $\boldsymbol{p}^{\mathrm{st}} \cdot \boldsymbol{u} = u^{\mathrm{st}}$ is the offset term every round, and thus $\pi(\boldsymbol{p}^{\mathrm{st}}, \mathbf{x}, \mathbf{y}) = 0$.

Definition 2 (normalization)**.** *We define the normalization function* $\mathrm{Norm} : \prod_{s \in \mathcal{S}} \mathbb{R}_+^m \mapsto \prod_{s \in \mathcal{S}} \mathrm{int}(\Delta^{m-1})$ *as*

$$\mathrm{Norm}(\mathbf{x}) = \left\{ \frac{x^{a|s}}{\sum_{a'} x^{a'|s}} \right\}_{a,s}, \tag{6}$$

In this definition, $\mathrm{Norm}(\mathbf{x})$ satisfies the condition of probability variables for all s.
Based on these definitions, we formulate discretized MMRD and MMGA as Algorithm 1 and 2.

Algorithm 1 Discretized MMRD

Input: η

1: **for** $t = 0, 1, 2, \cdots$ **do**
2: X chooses a with probability $x^{a|s_i}$
3: (Y chooses b with probability $y^{b|s_i}$)
4: $s_{i'} \leftarrow abs_i^-$
5: $x^{a|s_i} \leftarrow x^{a|s_i} + \eta \pi(\boldsymbol{e}_{i'}, \mathbf{x}, \mathbf{y})$
6: $\mathbf{x} \leftarrow \mathrm{Norm}(\mathbf{x})$
7: $s_i \leftarrow s_{i'}$
8: **end for**

Algorithm 1 (Discretized MMRD) takes its learning rate η as an input. In each time step, the players choose their actions following their strategies (lines 2 and 3), while the state is updated by their chosen actions (lines 4 and 7). Then, each player reinforces its strategy by how much payoff the chosen action brings up to the future. Here, note that for simplicity, this payoff is given by an expected payoff (line 5).

Algorithm 2 Discretized MMGA

Input: η, γ

1: **for** $t = 0, 1, 2, \cdots$ **do**
2: **for** $a \in \mathcal{A}$, $s \in \mathcal{S}$ **do**
3: $\mathbf{x}' \leftarrow \mathbf{x}$
4: $x'^{a|s} \leftarrow x'^{a|s} + \gamma$
5: $\Delta^{a|s} \leftarrow \dfrac{u^{\mathrm{st}}(\mathrm{Norm}(\mathbf{x}'), \mathbf{y}) - u^{\mathrm{st}}(\mathbf{x}, \mathbf{y})}{\gamma}$
6: **end for**
7: **for** $a \in \mathcal{A}$, $s \in \mathcal{S}$ **do**
8: $x^{a|s} \leftarrow x^{a|s}(1 + \eta \Delta^{a|s})$
9: **end for**
10: $\mathbf{x} \leftarrow \mathrm{Norm}(\mathbf{x})$
11: **end for**