Preliminary Draft: please do not cite or circulate.

Outcome Selection with Algorithmic Learners

Francesco Giordano [*] Mateus Hiro Nagata[†]

July 12, 2025

Abstract

We define a new procedure to nudge the selection of desirable outcomes in games played by algorithms. We consider the case where agents use a learning algorithm to play a repeated game. The innovative feature is to introduce a correlation device: decision makers update the values assigned to each action given the past actions performance and a payoff irrelevant message. Messages, which can be either public or private, are correlated among players. The probability distribution over messages is either fixed or time-varying according to some welfare criterion. We ask the following questions: do algorithms learn desirable correlated equilibria? Does information improves welfare and fairness when algorithms compete? We give a partial answer to the above questions based on simulations.

1 Introduction

People choose actions, observe outcomes, and evaluate whether their choice was good. Rinse and repeat. Can such trial-and-error procedure teach agents how to make strategic decisions to eventually converge to equilibrium play? This is the core question of the learning literature in game theory.

In this study, we introduce a method to augment existing algorithms to induce welfare-improving equilibria. The core idea is to allow players to condition their play on payoff-irrelevant signals, which we denote as messages - generated by a mediator. We focus on simultaneous-move repeated games, where agents repeatedly interact using reinforcement learning. In particular, we augment an algorithm that is important in many fields. In the computer science community, it is called Hedge, Multiplicative Weights Update or stateless Q-learning (Bailey and Piliouras, 2018; Cohen et al., 2017; Leonardos and Piliouras, 2022); in the economic literature, it is known as the Weighted Stochastic Fictitious Play in the economic literature (Pangallo et al., 2022) .

Several classes of existing learning algorithms ensure that the empirical distribution of play converges to the set of coarse correlated equilibria (Hart and Mas-Colell, 2000; Foster and Vohra, 1997), but not necessarily to a single (coarse) correlated equilibrium, so cycles and failing of coordination are potential outcomes of learning. Additionally, in the case of convergence to a specific equilibrium, often it may converge to undesirable, Pareto-dominated outcomes (Marden, 2017; Barman and Ligett, 2015).

Our paper addresses the question of whether it is possible to induce learning algorithms achieve welfare-superior correlated equilibrium (outside of the convex-hull of the Nash equilibria). Our findings show that through the introduction of private messages and a carefully designed information structure, agents converge to it with positive probability. This contrasts to other learning algorithms in the literature. As far as we know, this is the first instance of explicit integration of messages in the learning scheme outside the context of Markov games Greenwald et al. (2003).

A surprising result is that theoretically optimal information structure rarely induces learning algorithms to play the optimal correlated equilibrium. Oftentimes, it converges to the Nash equilibrium, which indicates that perhaps Nash is somehow more “stable”. On the other hand, a message distribution with stronger incentive compatibility induced correlated equilibrium much more often. If agents are to learn about the process, the correlation between messages and opponent’s actions, arguably, a message distribution that actually induces the correlated equilibrium

[*] HEC Paris, Economics and Decision Sciences, PhD Student

[†] HEC Paris, Economics and Decision Sciences, PhD Student

is superior. To capture this intuition, we define **Price of Learning**, which is the ratio between welfare loss incurred of learning the game, in contrast to agents that actually know all the primitives of the game, and the theoretically optimal social welfare.

This work is motivated by the use of learning algorithms as an approximation to the behavior of boundedly rational agents (Fudenberg and Levine, 1998). For instance, Erev and Roth (1998) and Goeree and Holt (2001) showed that algorithms that incorporate an exploration component and memory perform the best, which are the only parameters of our model. In particular, it is an instance of sunspot equilibria (Cass and Shell, 1983; Duffy and Fisher, 2005).

Our work builds upon the "steering problem" (Canyakmaz et al., 2024; Zhang et al., 2023). In this problem, which is a repeated game, a mediator, that does not know the player's underlying learning algorithm, but knows the potential payoffs of any action profile, tries to steer/induce/nudge players to asymptotically converge to play a desirable outcome. In this literature, the utility is decomposed by a base payoff that results from each player's action and a control signal that is defined by the mediator. The mediator approximates the underlying algorithm, predicts the mixed action that arises from the choice of control and choose the best control.

Our approach is inspired by the use of correlated information in repeated games: foundational contributions on correlation and communication in multistage games date back to Myerson (1986) and Forges (1988).

We implement the algorithm and test it on several benchmark classes of repeated simultaneous-move games. Simulation results confirm that, for appropriate parameterizations, the learning process reliably converges to desirable outcomes that improve upon the equilibria typically selected by standard methods. As a point of comparison, we use the Hedge algorithm as a baseline. However, our methodology is compatible with any independent learning algorithm, including policy gradient techniques. We conjecture that the insights obtained from our simulations generalize beyond the specific examples considered here.

The augmentation can be applied to a large class of myopic learning algorithms. In particular, we focus on the *Hedge algorithm* which is a generalization of fictitious play and best-response dynamics, but the class of myopic algorithms also include stimulus-response type of models. Forward-looking algorithms on the other hand, most notably Q-learning, became relevant in the economic literature (Calvano et al., 2020; Dolgopolov, 2024; Shoham et al., 2007). In general, myopic learning algorithms are used to study learning in one-shot games, while forward-looking algorithms are more cut of to repeated-games scenarios, especially collusive scenarios.

1.1 Outline

The paper is structured as follows. In Section 2, we first describe the baseline dynamics, then introduce conditional learning with messages under a stationary message distribution, and finally extend the framework to allow for adaptive message distributions and prove analytical results. In Section 3, we present a pseudocode implementation of our algorithm along with simulation results across various classes of games. In Section 4, we discuss reasonable extensions and questions that need further analysis in our framework. Finally, in Section 5, we conclude with a discussion and highlight the open research questions.

2 Dynamics

We focus on finite, repeated, simultaneous-move games with complete information described by a tuple $(N, \mathbf{A}, \{u_i\}_{i\in N})$, where N is the finite set of players, $\mathbf{A} = \times_i A_i$ is the finite set of action profiles and $u_i : \mathbf{A} \to \mathbb{R}$ is the utility function of player $i \in N$. The general learning dynamics we consider, Hedge, follows a stateless, discrete-time learning process. For any player and at any time period, each action $a_i \in A_i$ is associated with a $Q-$value, $Q^t_{a_i}$, which represents the quality or attraction the decision-maker assigns to his action. Pure action at time t, a^t_i, are chosen with probability $x^t_{a_i}$, which is defined by the softmax operation on the $Q-$values

$$x^t_{a_i} = \frac{\exp(\beta Q^t_{a_i})}{\sum_{a'_i \in A_i} \exp\left(\beta Q^t_{a'_i}\right)}$$

for a given parameter $\beta \in [0, +\infty)$. A higher value of β prioritizes exploitation over exploration. It also captures the idea that actions with higher $Q-$value are chosen more often, but the decision-maker may make errors or may not be confident on the assessment of how good an action is. The softmax choice rule is commonly applied in Discrete Choice analysis (Ben-Akiva and Lerman, 1985; Echenique and Saito, 2019) and satisfies Luce's Choice axioms (Luce et al., 1959), and is the basis for random utility models Anderson et al. (1992) thus being a reasonable way to make choices.

After the choice, the $Q-$values are updated according according to the following rule:

$$Q_{a_i}^{t+1} = (1-\alpha)Q_{a_i}^t + u_t(a_i, a_{-i}^t),$$

where a_{-i}^t indicates the action at time t of the opponents. In these expressions, $\alpha \in [0,1]$ is the memory-loss parameter that determines how much of the past $Q-$value is retained in the update.

2.1 Conditional Learning with Messages

We propose a modification of Hedge that incorporates a correlation device and message-dependent $Q-$values. This approach augments standard learning dynamics by introducing a finite set of messages, which can either be public or private. Let $(M_i)_{i\in N} = M$ represent the cartesian product of individual message sets $m_i \in M_i$. $\eta \in \Delta(M)$ representing the probability distribution of message $m = (m_i)_{i\in N}$ to be sampled at time t that is fixed and i.i.d. across time and independent of the past. This defines the game with messages (N, A, u, M). Both the $Q-$value and mixed strategy updates are performed conditionally on the private message. The learning dynamics remain as previously described, but is applied conditional on a message.

Specifically, at any time period, each action for any player is associated with a value that evolves according to the following process: at iteration t, a message $m = (m_i)_{i\in N}$ is drawn from $\eta \in \Delta(M)$ which in turn defines the mixed action:

$$x_{a_i}^t(m_i) = \mathbb{1}_{\{m_i^t = m_i\}} \frac{\exp\left(\beta Q_{a_i}^t(m_i)\right)}{\sum_{a_i' \in A_i} \exp\left(\beta Q_{a_i'}^t(m_i)\right)}.$$

which, for every fixed m_i, is a probability distribution over the actions of player i. For instance, the choice between taking an umbrella a_i or not a_i' on a cloudy day m_i depends only on the experiences associated with cloudy days. $Q-$value update applies only conditional on the message m_i

$$Q_{a_i}^{t+1}(m_i) = Q_{a_i}^t(m_i) + \mathbb{1}_{\{m_i^{t+1} = m_i\}}[u_i(a_i, a_{-i}^{t+1}) - \alpha Q_{a_i}^t(m_i)].$$

with initial $Q-$values $Q_{a_i}^0(m_i) = 0$ for all players, actions and messages, showing starting indifference or ignorance towards all actions.

In contrast to the baseline framework, there is now one Q-value for each pair of action and message, and hence the mixed strategies is defined for the pair. The resulting output is the pushforward distribution derived from the message distribution and the conditional mixed strategy. This yields a correlated strategy profile, which extends beyond what can be achieved by the baseline learning dynamics introduced previously, where players' behaviors remain independent. Notice that a specific case of this framework involves public messages, which effectively runs several reinforcement learning algorithms in parallel and randomizes the choice. An extension of this algorithm involves adapting over time the message distribution itself, allowing for time-varying distributions η^t and can be found on the Section 4.1.

The correlated equilibrium in the game with messages (N, A, u, M) can be defined as the collection of mixed actions $(x_{a_i}(m_i))_{a_i \in A_i, m_i \in M_i}$ and message distribution $\eta \in \Delta(M)$ such that for all agents i, all actions a_i and all messages m_i, the following inequality holds:

$$\sum_{a_{-i}\in A_{-i}} \sum_{m_{-i}\in M_{-i}} \eta(m_i, m_{-i}) x_{a_i}^t(m_i) x_{a_{-i}}^t(m_{-i})[u_i(a_i, a_{-i}) - u_i(a_i', a_{-i})] \geq 0. \tag{1}$$

We consider two primary measures for evaluating outcome desirability: social welfare and fairness. Social welfare is defined as

$$SW = \sum_i \sum_m \sum_a \eta(m) x_{a_i}^t(m) u_i(a, x_{a_{-i}}),$$

while fairness as

$$F = \min_i \sum_m \sum_a \eta(m) x_{a_i}^t(m) u_i(a, x_{a_{-i}}).$$

2.2 Analytical Results

In order to rigorously analyze the behavior of this learning scheme, let us study its **continuous-time approximation**. This approximation transforms the original stochastic learning problem into a deterministic dynamical system, making it more amenable to mathematical analysis. In particular, it enables a tractable study of fixed points, which can yield valuable insights into long-run behavior and convergence properties, serving as a benchmark for understanding the learning dynamics (Pangallo et al., 2022).

Naturally, one may ask whether this approximation faithfully reflects the behavior of the discrete-time system. There are two main approaches to mitigate this problem. One is empirical validation through simulations - the approach taken in this paper. The other one is to use stochastic approximation theory - which typically involves the use of time-varying parameters.

We now define the **continuous time** counterpart of the Hedge $Q-$value. Consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}^t)_{t\geq 0}, \mathbb{P})$, where $\mathcal{F}^t$ encodes the history of play up to time t. The $Q-$values are treated as stochastic processes adapted to this filtration.

The continuous-time evolution of $Q-$value for player i, action a_i, private message $m_i \in M_i$, time t is defined via the expected infinitesimal increment:

$$\dot{Q}^t_{a_i}(m_i) = \frac{d}{dt}\mathbb{E}[Q^t_{a_i}(m_i) \mid \mathcal{F}^t] = \lim_{\delta\to 0^+} \frac{\mathbb{E}_{m\sim\eta(\cdot),a_{-i}\sim x^{t+\delta}_{a_{-i}}(m_{-i})}[Q^{t+\delta}_{a_i}(m_i) - Q^t_{a_i}(m_i) \mid \mathcal{F}^t]}{\delta} \tag{2}$$

This defines a deterministic approximation of the $Q-$value dynamics. The expectation is taken over the joint distribution of messages $m = (m_i, m_{-i}) \sim \eta$, and the opponent's actions $a^{t+\delta}_{-i} \sim x^{t+\delta}_{a_{-i}}(m_{-i})$. Since updates occur after the realization, $x^{t+\delta}_{a_{-i}}(m_{-i})$ is not treated as a random variable, but as a deterministic quantity at time $t+\delta$.

We define the expected utility of playing action $a_i \in A_i$ given message $m_i \in M_i$ as:

$$u^t_i(a_i|m_i) = \sum_{a_{-i}}\sum_{m_{-i}} u_i(a_i, a_{-i})x^t_{a_{-i}}(m_{-i})\eta(m_{-i}|m_i). \tag{3}$$

where $\eta(m_{-i}|m_i)$ represents the conditional probability of the opponent's message being m_{-i}, given that i's message is m_i. If the messages are public, then this would be a degenerate probability distribution and if the messages were independent, it would simply be $\eta(m_{-i})$.

Now, the Hedge algorithm can be analyzed by the system of deterministic ODEs defined in Lemma 1.

Lemma 1. *The correlated Hedge algorithm has the following continuous-time equivalent:*

$$\dot{Q}^t_{a_i}(m_i) = \eta(m_i)[u^t_i(a_i|m_i) - \alpha Q^t_{a_i}(m_i)] \tag{4}$$

$$\dot{x}^t_{a_i}(m_i) = \beta x^t_{a_i}(m_i)\left[\dot{Q}^t_{a_i}(m_i) - \sum_{a'\in A_i}\dot{Q}^t_{a'_i}(m_i)x^t_{a'_i}(m_i)\right], \tag{5}$$

Proof. We prove in two parts: first for the $Q-$value dynamics, then for the mixed action dynamics.

$Q-$**value dynamics**: the average Q-value update for a small step size $\delta > 0$ is

$$\frac{\mathbb{E}[Q^{t+\delta}_{a_i}(m_i) - Q^t_{a_i}(m_i)|\mathcal{F}^t]}{\delta} = \frac{\delta\mathbb{E}[\mathbb{1}(m^{t+\delta}_i = m_i)(u_i(a_i, a^{t+\delta}_{-i}) - \alpha Q^t_{a_i}(m_i))]}{\delta}$$

Now, evaluating the expected value:

$$\begin{aligned}\mathbb{E}[\mathbb{1}(m^{t+\delta}_i = m_i)u_i(a_i, a^{t+\delta}_{-i})] &= \eta(m_i)E[u_i(a_i, a^{t+\delta}_{-i})|m^{t+\delta}_i = m_i] \\ \eta(m_i)\sum_{m_{-i}}\eta(m_{-i} \mid m_i)\sum_{a_{-i}}u_i(a_i, a_{-i})x^{t+\delta}_{a_{-i}}(m_{-i}) &= \eta(m_i)u^{t+\delta}_i(a_i \mid m_i) \\ &= \eta(m_i)u^{t+\delta}_i(a_i|m_i)\end{aligned}$$

Therefore,

$$\begin{aligned}\lim_{\delta\to 0^+}\frac{\mathbb{E}[Q^{t+\delta}_{a_i}(m_i)-Q^t_{a_i}(m_i)|\mathcal{F}^t]}{\delta} &= \lim_{\delta\to 0^+}\eta(m_i)[\mathbb{E}[u_i(a_i,a^{t+\delta}_{-i})|m^{t+\delta}_i=m_i,\mathcal{F}^t]-\alpha Q^t_{a_i}(m_i)] \\ &= \eta(m_i)[u^t_i(a_i|m_i)-\alpha Q^t_{a_i}(m_i)].\end{aligned}$$

Mixed strategy dynamics:

$$\begin{aligned}\dot{x}^t_{a_i}(m_i) &= \frac{d}{dt}\frac{\exp(\beta Q^t_{a_i}(m_i))}{\sum_{a'_i\in A_i}\exp(\beta Q^t_{a'_i}(m_i))} \\ &= \frac{\beta\dot{Q}^t_{a_i}(m_i)\exp(\beta Q^t_{a_i}(m_i))\left(\sum_{a'_i\in A_i}\exp(\beta Q^t_{a'_i}(m_i))\right)-\exp(\beta Q^t_{a_i}(m_i))\left(\sum_{a'_i\in A_i}\beta\dot{Q}^t_{a'_i}(m_i)\exp(\beta Q^t_{a'_i}(m_i))\right)}{\left(\sum_{a'_i\in A_i}\exp(\beta Q^t_{a'_i}(m_i))\right)^2}\end{aligned}$$

Which results in

$$\dot{x}^t_{a_i}(m_i)=\beta x^t_{a_i}(m_i)\left[\dot{Q}^t_{a_i}(m_i)-\sum_{a'\in A_i}\dot{Q}^t_{a'_i}(m_i)x^t_{a'_i}(m_i)\right].$$

□

Now, the formula 5 is a Replicator Equation Schuster and Sigmund (1983), a generalization of Replicator Dynamics Hofbauer and Sigmund (2003). Basically, it states that the evolution of the continuous-time equivalent grows proportional to the difference between the evolution of $Q-$value towards action a_i given message m_i and the average evolution of $Q-$values given messages.

In this context, β and $x^t_{a_i}(m_i)$ both amplify this effect. Specifically, if at time t, given message m_i, the $Q-$value for action a_i is very high yet the utility of period t was very low, the agent is willing to correct the mixed action a lot. On the other hand, if $x^t_{a_i}(m_i)$ was very low with respect to the average, it expresses careful updating.

To see this clearly, we can decompose the origins of the evolution of mixed action into two different factors. The first is the difference between the expected utility of playing a_i given m_i and the utility of playing the current mixed action. The second is related to the memory and has a more difficult interpretation.

Lemma 2. *The continuous-time correlated hedge can be alternatively described by the formula*

$$\dot{x}^t_{a_i}(m_i)=\beta x^t_{a_i}(m_i)\eta(m_i)\left[u^t_i(a_i|m_i)-\sum_{a'_i\in A_i}x^t_{a'_i}(m_i)u^t_i(a'_i|m_i)\right]- \tag{6}$$

$$\alpha x^t_{a_i}(m_i)\eta(m_i)\left[\ln(x^t_{a_i}(m_i))-\sum_{a'_i\in A_i}x^t_{a'_i}(m_i)\ln(x^t_{a'_i}(m_i))\right] \tag{7}$$

Proof. We can rewrite the mixed action as

$$\ln(x^t_{a_i}(m_i))=\beta Q^t_{a_i}(m_i)-\ln\left(\sum_{a'_i\in A_i}e^{\beta Q^t_{a'_i}(m)}\right)$$

Rearranging

$$Q^t_{a_i}(m_i)=\frac{1}{\beta}\ln\left(x^t_{a_i}(m_i)\right)+\frac{1}{\beta}\ln\left(\sum_{a'_i\in A_i}e^{\beta Q^t_{a'_i}(m)}\right).$$

Thus, arriving at:

$$\dot{x}^t_{a_i}(m_i) = \beta x^t_{a_i}(m_i)\eta(m_i)\left[u^t_i(a_i|m_i) - \sum_{a_i \in A_i} x^t_{a_i}(m_i)u^t_i(a_i|m_i)\right] -$$
$$\alpha x^t_{a_i}(m_i)\eta(m_i)\left[\ln(x^t_{a_i}(m_i)) - \sum_{a'_i \in A_i} x^t_{a'_i}(m_i)\ln(x^t_{a'_i}(m_i))\right]$$

□

Observation 1. *The first expression in brackets is positive if and only if the action a_i's average utility is higher than the utility of playing the mixed action $x^t_{a_i}$. Let us denote this as the **reinforcement condition***

$$\beta x^t_{a_i}(m_i)\eta(m_i)\left[u^t_i(a_i|m_i) - \sum_{a'_i \in A_i} x^t_{a'_i}(m_i)u^t_i(a'_i|m_i)\right].$$

Furthermore, β only influences the system through the reinforcement condition.

Now, to investigate further, we analyze the stability of the fixed points of the system and its relations with correlated equilibrium.

Definition 1. *A fixed point $x^* = (x^*_{a_i}(m_i))_{a_i \in A_i, m_i \in M_i, i \in N}$ is **stable** if for every $\varepsilon > 0$, there is a $\delta > 0$ such that*

$$||x^0 - x^*|| \leq \delta \Rightarrow ||x^t - x^*|| \leq \varepsilon$$

where $x^t = (x^t_{a_i}(m_i))_{a_i \in A_i, m_i \in M_i, i \in N}$.

Definition 2. *A fixed point $x^* = (x^*_{a_i}(m_i))_{a_i \in A_i, m_i \in M_i, i \in N}$ is **asymptotically stable** if there is a $\delta > 0$ such that*

$$||x^0 - x^*|| \leq \delta \Rightarrow \lim_{t \to \infty} x^t = x^*.$$

Now, let us describe Proposition 1, the main one in this paper. This proposition showcases the fixed points of the system.

Proposition 1. *Suppose a 2×2 game with 2 messages for each player playing Correlated Hedge. Then, all pure strategy profiles given a message $(x^t_{a_i}(m_i))_{m_i \in M_i, a_i \in A_i, i \in N} \in \{0,1\}^{|M| \times |A|}$, are fixed points. Furthermore, suppose full-memory, $\alpha = 0$, then a pure strategy profile is stable if and only if it is a correlated equilibrium.*

I have proved that all pure actions given messages are fixed points, but I am re-evaluating the final step of the proof. I write down the reasoning
The proof uses the convention $\lim_{\gamma \to 0^+} \gamma \ln(\gamma) = 0\ln(0) = 0$.

Proof. Suppose $x^t_{a_i}(m_i) = 0$, then

$$\dot{x}^t_{a_i}(m_i) = 0\ln(0).$$

Suppose $x^t_{a_i}(m_i) = 1$, then

$$\dot{x}^t_{a_i}(m_i) = \beta\eta(m_i)\left[u^t_i(a_i \mid m_i) - u^t_i(a_i \mid m_i)\right] - \alpha\eta(m_i)\left[\ln(1) - \ln(1)\right] = 0.$$

Concerning stability, one way to prove stability is through **Lyapunov Linearization Theorem** which states that if all eigenvalues of the Jacobian have strictily negative parts, then it is asymptotically stable Hirsch et al. (2013). The rest of the proof is very algebraic, and complicated but it simplifies very nicely. So, proceed with the idea that the expressions are complicated but I am just writing the Jacobian that reduces nicely.

Since we are in a 2×2 game, we could write the continuous-time equivalent as

$$\dot{x}^t_{a_i}(m_i) = \beta x^t_{a_i}(m_i)\eta(m_i)(1 - x^t_{a_i}(m_i))\left[u^t_i(a_i|m_i) - u^t_i(a'_i|m_i)\right] -$$
$$\alpha x^t_{a_i}(m_i)\eta(m_i)(1 - x^t_{a_i}(m_i))\left[\ln\left(\frac{x^t_{a_i}(m_i)}{1 - x^t_{a_i}(m_i)}\right)\right]$$

and the average utility of playing a_i given m_i as

$$\begin{aligned} u_i^t(a_i|m_i) = \eta(m_{-i}|m_i)[u_i(a_i, a_{-i})x^t_{a_{-i}}(m_{-i}) + u_i(a_i, a'_{-i})(1 - x^t_{a_{-i}}(m_{-i}))] + \\ \eta(m'_{-i}|m_i)[u_i(a_i, a_{-i})(1 - x^t_{a'_{-i}}(m'_{-i})) + u_i(a_i, a'_{-i})x_{a'_{-i}}(m'_{-i}))]. \end{aligned}$$

The Jacobian is defined as follows:

$$J(x) = \begin{bmatrix} \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a'_{-i}}(m'_{-i})} \\ \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_{-i}}(m'_{-i})} \\ \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a'_{-i}}(m'_{-i})} \\ \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a'_{-i}}(m'_{-i})} \end{bmatrix}.$$

The partials are on the appendix 9.

The Jacobian is defined as follows:

$$J(x) = \begin{bmatrix} \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a'_{-i}}(m'_{-i})} \\ \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_{-i}}(m'_{-i})} \\ \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a'_{-i}}(m'_{-i})} \\ \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a'_{-i}}(m'_{-i})} \end{bmatrix}.$$

It follows that:

$$J(x) = \begin{bmatrix} \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_i}(m_i)} & 0 & \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a'_{-i}}(m'_{-i})} \\ 0 & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_{-i}}(m'_{-i})} \\ \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a_{-i}}(m_{-i})} & 0 \\ \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a'_i}(m'_i)} & 0 & \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a'_{-i}}(m'_{-i})} \end{bmatrix}.$$

To compute the eigenvalues, we divide the Jacobian in

$$J(x) = \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

Hence, the eigenvalues are defined by the formula (Schur's formula)

$$\det(J(x) - \lambda I) = \det(A - \lambda I_2) \cdot \det(D - \lambda I_2 - C(A - \lambda I_2)^{-1}B)$$

$$\det(A - \lambda I_2) = \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_i}(m_i)} \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_i}(m'_i)} - \lambda \left(\frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_i}(m_i)} + \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_i}(m'_i)} \right) + \lambda^2$$

Also, we have

$$(A - \lambda I_2)^{-1} = \frac{1}{\left| \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_i}(m_i)} \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_i}(m'_i)} - \lambda \left(\frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_i}(m_i)} + \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_i}(m'_i)} \right) + \lambda^2 \right|} \begin{bmatrix} \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_i}(m'_i)} & 0 \\ 0 & \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_i}(m_i)} \end{bmatrix}$$

[HERE]

$$J(x) = \begin{bmatrix} \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_i}(m_i)} & 0 & \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a'_{-i}}(m'_{-i})} \\ 0 & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_{-i}}(m'_{-i})} \\ \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a_{-i}}(m_{-i})} & 0 \\ \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a'_i}(m'_i)} & 0 & \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a'_{-i}}(m'_{-i})} \end{bmatrix}$$

We write this as a block matrix:

$$J = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where:

$$A = \begin{bmatrix} a & 0 \\ 0 & a' \end{bmatrix}, \quad D = \begin{bmatrix} d & 0 \\ 0 & d' \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}, \quad C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

We compute the characteristic polynomial using the Schur complement:

$$\chi(\lambda) = \det(J - \lambda I_2) = \det(A_\lambda) \cdot \det\left(D_\lambda - C A_\lambda^{-1} B\right)$$

Let:

$$A_\lambda = A - \lambda I = \begin{bmatrix} a - \lambda & 0 \\ 0 & a' - \lambda \end{bmatrix}, \quad D_\lambda = D - \lambda I = \begin{bmatrix} d - \lambda & 0 \\ 0 & d' - \lambda \end{bmatrix}, \quad D_\lambda^{-1} = \begin{bmatrix} \frac{1}{d-\lambda} & 0 \\ 0 & \frac{1}{d'-\lambda} \end{bmatrix}$$

Now, calculating, we have

$$A_\lambda = \begin{bmatrix} a - \lambda & 0 \\ 0 & a' - \lambda \end{bmatrix}, \quad A_\lambda^{-1} = \begin{bmatrix} \frac{1}{a-\lambda} & 0 \\ 0 & \frac{1}{a'-\lambda} \end{bmatrix}$$

$$B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}, \quad C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

First, compute the intermediate product:

$$C A_\lambda^{-1} = \begin{bmatrix} \frac{c_{11}}{a-\lambda} & \frac{c_{12}}{a'-\lambda} \\ \frac{c_{21}}{a-\lambda} & \frac{c_{22}}{a'-\lambda} \end{bmatrix}$$

Now compute the final product:

$$C A_\lambda^{-1} B = \begin{bmatrix} \dfrac{c_{11} b_{11}}{a - \lambda} + \dfrac{c_{12} b_{21}}{a' - \lambda} & \dfrac{c_{11} b_{12}}{a - \lambda} + \dfrac{c_{12} b_{22}}{a' - \lambda} \\ \dfrac{c_{21} b_{11}}{a - \lambda} + \dfrac{c_{22} b_{21}}{a' - \lambda} & \dfrac{c_{21} b_{12}}{a - \lambda} + \dfrac{c_{22} b_{22}}{a' - \lambda} \end{bmatrix}$$

$$\det(C A_\lambda^{-1} B) = \left(\frac{c_{11} b_{11}}{a - \lambda} + \frac{c_{12} b_{21}}{a' - \lambda} \right) \left(\frac{c_{21} b_{12}}{a - \lambda} + \frac{c_{22} b_{22}}{a' - \lambda} \right) - \left(\frac{c_{11} b_{12}}{a - \lambda} + \frac{c_{12} b_{22}}{a' - \lambda} \right) \left(\frac{c_{21} b_{11}}{a - \lambda} + \frac{c_{22} b_{21}}{a' - \lambda} \right)$$

So, our determinants are

$$(a - \lambda)(a' - \lambda) \cdot \left(\left(\frac{c_{11} b_{11}}{a - \lambda} + \frac{c_{12} b_{21}}{a' - \lambda} \right) \left(\frac{c_{21} b_{12}}{a - \lambda} + \frac{c_{22} b_{22}}{a' - \lambda} \right) - \left(\frac{c_{11} b_{12}}{a - \lambda} + \frac{c_{12} b_{22}}{a' - \lambda} \right) \left(\frac{c_{21} b_{11}}{a - \lambda} + \frac{c_{22} b_{21}}{a' - \lambda} \right) \right)$$

Which simplifies to
Let $\alpha = a - \lambda$, $\beta = a' - \lambda$. Then:

$$\alpha\beta \cdot \left[\left(\frac{c_{11} b_{11}}{\alpha} + \frac{c_{12} b_{21}}{\beta} \right) \left(\frac{c_{21} b_{12}}{\alpha} + \frac{c_{22} b_{22}}{\beta} \right) - \left(\frac{c_{11} b_{12}}{\alpha} + \frac{c_{12} b_{22}}{\beta} \right) \left(\frac{c_{21} b_{11}}{\alpha} + \frac{c_{22} b_{21}}{\beta} \right) \right]$$

Expanding, we get:

$$= \beta(c_{11}b_{11}c_{21}b_{12}-c_{11}b_{12}c_{21}b_{11})+(c_{11}b_{11}c_{22}b_{22}+c_{12}b_{21}c_{21}b_{12}-c_{11}b_{12}c_{22}b_{21}-c_{12}b_{22}c_{21}b_{11})+\alpha(c_{12}b_{21}c_{22}b_{22}-c_{12}b_{22}c_{22}b_{21})$$

Therefore, the final result is:

$$(a-\lambda)(a'-\lambda)\cdot\det(CA_\lambda^{-1}B) = \beta(c_{11}b_{11}c_{21}b_{12}-c_{11}b_{12}c_{21}b_{11})+(c_{11}b_{11}c_{22}b_{22}+c_{12}b_{21}c_{21}b_{12}-c_{11}b_{12}c_{22}b_{21}-c_{12}b_{22}c_{21}b_{11})+\alpha(c_{12}b_{21}c_{22}$$

Then the Schur complement is:

$$S(\lambda) = A_\lambda - BD_\lambda^{-1}C$$

NOW: if pure-strategy given message, then only the self-interaction term is non-null, so the Jacobian can be re-written as:

$$J(x) = \begin{bmatrix} \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_i}(m_i)} & 0 & 0 & 0 \\ 0 & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_i}(m'_i)} & 0 & 0 \\ 0 & 0 & \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a_{-i}}(m_{-i})} & 0 \\ 0 & 0 & 0 & \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a'_{-i}}(m'_{-i})} \end{bmatrix}$$

Which means they are the eigenvalues. Which means that their real parts (they are purely real numbers) is negative iff

$$u_i^t(a_i|m_i) - u_i^t(a'_i|m_i) \leq 0$$

NOW,

$$CA_\lambda^{-1}B = \begin{bmatrix} \dfrac{c_{11}b_{11}}{a-\lambda} + \dfrac{c_{12}b_{21}}{a'-\lambda} & \dfrac{c_{11}b_{12}}{a-\lambda} + \dfrac{c_{12}b_{22}}{a'-\lambda} \\ \dfrac{c_{21}b_{11}}{a-\lambda} + \dfrac{c_{22}b_{21}}{a'-\lambda} & \dfrac{c_{21}b_{12}}{a-\lambda} + \dfrac{c_{22}b_{22}}{a'-\lambda} \end{bmatrix}$$

$$D_\lambda - CA_\lambda^{-1}B = \begin{bmatrix} (d-\lambda) - \left(\frac{c_{11}b_{11}}{a-\lambda} + \frac{c_{12}b_{21}}{a'-\lambda}\right) & -\left(\frac{c_{11}b_{12}}{a-\lambda} + \frac{c_{12}b_{22}}{a'-\lambda}\right) \\ -\left(\frac{c_{21}b_{11}}{a-\lambda} + \frac{c_{22}b_{21}}{a'-\lambda}\right) & (d'-\lambda) - \left(\frac{c_{21}b_{12}}{a-\lambda} + \frac{c_{22}b_{22}}{a'-\lambda}\right) \end{bmatrix}$$

$$\det(D_\lambda - CA_\lambda^{-1}B) = (d-\lambda)(d'-\lambda) - (d'-\lambda)\left(\frac{c_{11}b_{11}}{a-\lambda} + \frac{c_{12}b_{21}}{a'-\lambda}\right)$$
$$-(d-\lambda)\left(\frac{c_{21}b_{12}}{a-\lambda} + \frac{c_{22}b_{22}}{a'-\lambda}\right) + \frac{((a'-\lambda)c_{11}b_{11} + (a-\lambda)c_{12}b_{12})\,((a'-\lambda)c_{21}b_{12} + (a-\lambda)c_{22}b_{22})}{(a-\lambda)(a'-\lambda)}$$

Therefore,

$$\det(D_\lambda - CA_\lambda^{-1}B) = (d-\lambda)(d'-\lambda) - (d-\lambda)\left(\frac{c_{21}b_{12}}{a-\lambda} + \frac{c_{22}b_{22}}{a'-\lambda}\right) - (d'-\lambda)\left(\frac{c_{11}b_{11}}{a-\lambda} + \frac{c_{12}b_{21}}{a'-\lambda}\right) + \frac{\det(C)\det(B)}{(a-\lambda)(a'-\lambda)}$$

$$\det(A_\lambda)\det(D_\lambda - CA_\lambda^{-1}B) = (a-\lambda)(a'-\lambda)(d-\lambda)(d'-\lambda) - (d'-\lambda)\left((a'-\lambda)c_{11}b_{11} + (a-\lambda)c_{12}b_{21}\right)$$
$$-(d-\lambda)\left((a'-\lambda)c_{21}b_{12} + (a-\lambda)c_{22}b_{22}\right) + \left((a'-\lambda)c_{11}b_{11} + (a-\lambda)c_{12}b_{12}\right)\left((a'-\lambda)c_{21}b_{12} + (a-\lambda)c_{22}b_{22}\right)$$

The final step that needs validation is the evaluation of the Jacobian. I believe that (asymptotic) stability would be equivalent to correlated equilibria when $\alpha = 0$, given similar results when the algorithm has no messages Pangallo et al. (2022). A correlated equilibrium, characterized by $\forall i \in N, \forall a_i, a'_i \in A_i, \forall m_i \in M_i$:

$$\sum_{a_{-i}\in A_{-i}}\sum_{m_{-i}\in M_{-i}}\eta(m_i,m_{-i})x^t_{a_i}(m_i)x^t_{a_{-i}}(m_{-i})[u_i(a_i,a_{-i})-u_i(a'_i,a_{-i})]\geq 0. \tag{8}$$

This is equivalent to

$$\sum_{a_{-i}\in A_{-i}}\sum_{m_{-i}\in M_{-i}}\eta(m_i|m_{-i})\eta(m_{-i})x^t_{a_i}(m_i)x^t_{a_{-i}}(m_{-i})[u_i(a_i,a_{-i})-u^t_i(a'_i,a_{-i})]\geq 0$$

$$\eta(m_i)x^t_{a_i}(m_i)[u^t_i(a_i\mid m_i)-u^t_i(a'_i\mid m_i)]\geq 0$$

□

Finally, there are some results in the literature pointing to the equivalence between interior fixed points and quantal response equilibria (Leonardos and Piliouras, 2022; Pangallo et al., 2022). A correspondent solution concept for the game with messages is called the Quantal Correlated Equilibrium (QCE), specifically the per-signal Quantal Correlated Equilibrium (S-QCE) (Černỳ et al., 2022). It is defined in the signaling game (N, A, u, M) as the signaling scheme $\eta \in \Delta(M)$ and mixed action $x_{a_i}(m_i)$ if there is a function $q_i(\cdot)$ that satisfies

$$x_{a_i}(m_i)=\frac{q_i(u_i(a_i|m_i))}{\sum_{a'_i\in A_i}q_i(u_i(a'_i|m_i))}$$

and $q_i(\cdot)$ is a strictly positive and increasing function. In particular, we prove that it follows for $q_i(z)=\exp(\frac{\beta}{\alpha}z)$ in Proposition 2.

Proposition 2. *Suppose a 2×2 game with 2 messages for each player playing Correlated Hedge. Then, all fixed points are equivalent to a Per-signal Quantal Correlated Equilibrium (S-QCE).*

Proof. The fixed-point condition in Lemma 2 can be expressed as:

$$\beta\left[u^t_i(a_i|m_i)-\sum_{a'_i\in A_i}x^t_{a'_i}(m_i)u^t_i(a'_i|m_i)\right]=\alpha\left[\ln(x^t_{a_i}(m_i))-\sum_{a'_i\in A_i}x^t_{a'_i}(m_i)\ln(x^t_{a'_i}(m_i))\right]$$

Since we restrict our attention to 2×2 games, we have that

$$\beta\left[(1-x^t_{a_i}(m_i))u^t_i(a_i|m_i)-(1-x^t_{a_i}(m_i))u^t_i(a'_i|m_i)\right]=\alpha\left[(1-x^t_{a_i}(m_i))\ln(x^t_{a_i}(m_i))-(1-x^t_{a_i})(m_i)\ln(x^t_{a'_i}(m_i))\right]$$

$$\beta\left[u^t_i(a_i|m_i)-u^t_i(a'_i|m_i)\right]=\alpha\left[\ln(x^t_{a_i}(m_i))-\ln(x^t_{a'_i}(m_i))\right]$$

by the properties of logarithm, we have

$$\frac{\beta}{\alpha}\left[u^t_i(a_i|m_i)-u^t_i(a'_i|m_i)\right]=\left[\ln\left(\frac{x^t_{a_i}(m_i)}{x^t_{a'_i}(m_i)}\right)\right].$$

Since $1-x^t_{a_i}(m_i)=x^t_{a'_i}(m_i)$, it follows that

$$\frac{\beta}{\alpha}\left[u^t_i(a_i|m_i)-u^t_i(a'_i|m_i)\right]=\left[\ln\left(\frac{x^t_{a_i}(m_i)}{1-x^t_{a_i}(m_i)}\right)\right]$$

$$\left(\frac{x^t_{a_i}(m_i)}{1-x^t_{a_i}(m_i)}\right)=e^{\frac{\beta}{\alpha}\left[u^t_i(a_i|m_i)-u^t_i(a'_i|m_i)\right]}$$

$$\left(x^t_{a_i}(m_i)\right)=e^{\frac{\beta}{\alpha}\left[u^t_i(a_i|m_i)-u^t_i(a'_i|m_i)\right]}(1-x^t_{a_i}(m_i))$$

$$\left(x^t_{a_i}(m_i)\right)(1+e^{\frac{\beta}{\alpha}\left[u^t_i(a_i|m_i)-u^t_i(a'_i|m_i)\right]})=e^{\frac{\beta}{\alpha}\left[u^t_i(a_i|m_i)-u^t_i(a'_i|m_i)\right]}$$

$$x^t_{a_i}(m_i) = \frac{e^{\frac{\beta}{\alpha}\left[u^t_i(a_i|m_i) - u^t_i(a'_i|m_i)\right]}}{1 + e^{\frac{\beta}{\alpha}\left[u^t_i(a_i|m_i) - u^t_i(a'_i|m_i)\right]}}$$

Multiplying the right-hand-side numerator and denominator by the same factor, we have

$$x^t_{a_i}(m_i) = \frac{e^{\frac{\beta}{\alpha}\left[u^t_i(a_i|m_i)\right]}}{e^{\frac{\beta}{\alpha}[u^t_i(a'_i|m_i)]} + e^{\frac{\beta}{\alpha}\left[u^t_i(a_i|m_i)\right]}},$$

which is the formula for S-QCE when $q_i(z) = \exp(\frac{\beta}{\alpha}z)$. □

3 Results

We focus on the implementations of the above algorithm for the Chicken game (Aumann, 1987). The implementation with stationary message distribution and full feedback is presented in the pseudo-code Algorithm 1 that outputs the realized history of social welfare, fairness, and the realized frequency of play.

Algorithm 1 Algorithm with stationary private messages and full-feedback

```
Require:
  Finite game $(N, A, u)$
  Parameters $\alpha, \beta$
  Set $M_i = \{m_1, ..., m_n\}, i \in N, M = \times_{i\in N} M_i$ and probability distribution $\eta \in \Delta(M)$
  –
  Initialize $Q^t_{a_i}(m) = 0$ for all players $i \in N$, for all messages $m \in M$

  for $t = 1, \dots, T$ do                                   ▷ Simultaneously for all players $i$
      Draw $m \sim \eta$                                      ▷ Sample message from $\eta$
      $x^t_{a_i}(m_i) \leftarrow \Lambda(Q^t_i(m_i))$
      Draw $a^t_i \sim x^t_i(m)$                              ▷ Sample action from mixed strategy
      Compute $r^i(t) = [u_i(a_i, a^t_{-i})]_{a_i}$           ▷ Full feedback over actions
      $Q^{t+1}_i(m) \leftarrow (1 - \alpha)Q^t_i(m) + r^t_i$
  end for

  Output:
  Frequency of play, fairness, social welfare
```

We simulate the interaction between two *Correlated Hedge* algorithms with identical parameters (α, β) playing the Hawk-dove game for $T = 500$ rounds. To account for randomness, we run 100 independent simulations and study the last iterate mixed action of each run $x^T_{a_i}(m_i)$ and examine whether the algorithm reaches obedience (hence the welfare-improving correlated equilibrium) the Nash equilibrium or some other outcome. Since the softmax never exactly reaches 0 or 1, we use a 99.5% threshold to determine convergence.

Example 1. Hawk-Dove game with full feedback Consider the Hawk-Dove game with the following payoff structure

	a_2	b_2
a_1	$6, 6$	$2, 7$
b_1	$7, 2$	$0, 0$

Define the message sets $M_i = \{m_{a_i}, m_{b_i}\}, i = 1, 2$ with the joint messages (m_1, m_2) following the probability distribution:

$$\eta(m_{a_1}, m_{a_2}) = \eta(m_{a_1}, m_{b_2}) = \eta(m_{b_1}, m_{a_2}) = \frac{1}{3}.$$

The Hawk-dove game is the quintessential example to test our algorithm. By the revelation principle, a correlated equilibrium can be interpreted as obedience to the recommendations of an incentive-compatible correlation device. Hence, let us define $(x^T_{a_1}(m_{a_1}), x^T_{b_1}(m_{b_1}), x^T_{a_2}(m_{a_2}), x^T_{b_2}(m_{b_2})) = (1,1,1,1)$ as the **obedient strategy profile**, which will be the focus of our analysis.

The pure Nash equilibrium strategy profiles of the game are $(A_1, B_2), (A_2, B_1)$. From a social-welfare maximizer information-designer perspective, recommending (b_1, b_2) is not interesting. On the other hand, recommending (a_1, a_2) as long as it is incentive-compatible is welfare maximizing. Hence, let us compare the previous probability distribution of messages to the welfare-maximizing message distribution:

$$\eta'(m_{a_1}, m_{a_2}) = \frac{1}{2}, \eta'(m_{a_1}, m_{b_2}) = \eta'(m_{b_1}, m_{a_2}) = \frac{1}{4}.$$

The social welfare that results from obedience to the message distribution η is $SW_\eta = 10$, and to the message distribution η' is $SW_{\eta'} = 10.5$.

Nonetheless, the system may not converge to obedience. Learning algorithms reportedly converge to Pareto-worse equilibria (Fudenberg and Levine, 1998), may have chaotic behavior with no convergence (Sato et al., 2002; Galla and Farmer, 2013). Hence, η may induce higher social welfare if obedience is not satisfied with η'. This is what we found in our simulations.

Proposition 1 guarantees the equivalence of an obedient correlated equilibrium and an asymptotically stable fixed point in the continuous approximation. In other words, if the mixed action is close enough to the pure action correlated equilibrium, the system converges back to the fixed point. This is a local property. However, our simulations start with $Q-$values at 0 and hence the mixed actions start uniformly random. This means that convergence is not guaranteed and may depend on the specific message and pure action sampled in a given simulation, which is inherently stochastic.

Thus, we study the proportion of the simulations that converge to obedience.

The main result in the heatmaps in Figure 1. Given the baseline information distribution η, It displays the proportion, out of 100 simulations, of last-iterate convergence to **obedience** (on the left) and to **obedience or Nash equilibrium** on the right for different combinations of α, β. The grid divides the $\alpha \in [0,1]$ parameter space in 11 different values $\{0, 0.1, \ldots, 1.0\}$ and $\beta \in [0, 10]$ in 100 different values.

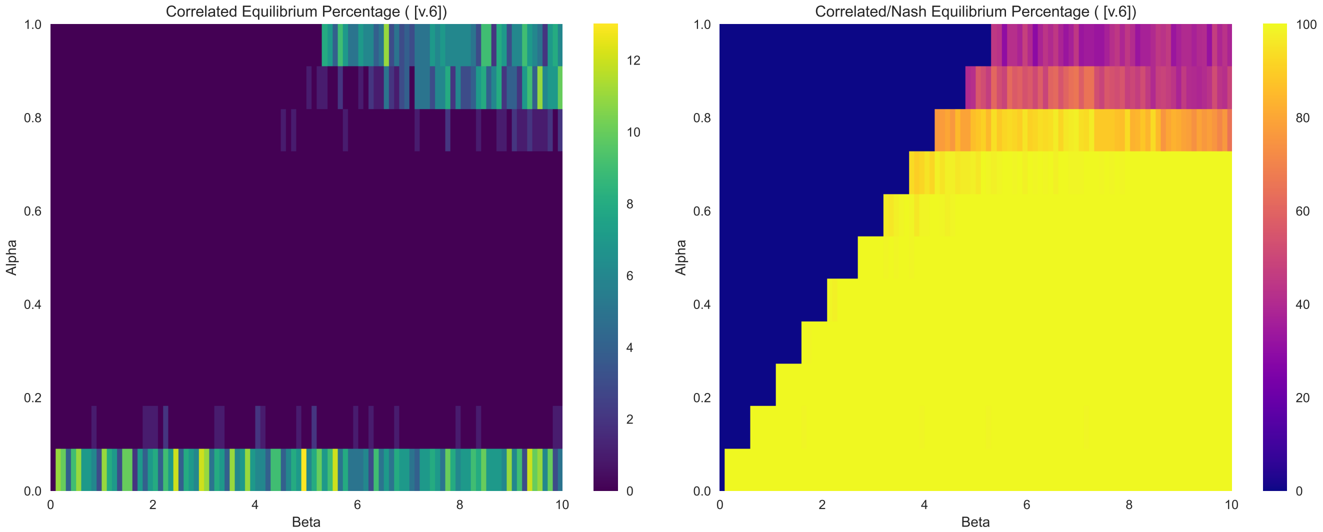


Figure 1: Heatmap of The Correlated Equilibrium Percentage (Excluding Nash Equilibrium) (Left) and Correlated Equilibrium Percentage (Right)

The graphs reveal an unexpected result: the obedient correlated equilibrium scenario emerges when the memory-loss parameter is $\alpha = 0$ and close to 1.

In the full-memory case, $\alpha = 0$, regardless of β, the algorithm converges to obedience around 10% of the time. In all cases, it converges to obedience or Pure Nash 100% of the time. Pure Nash equilibrium, in this case, refers to playing (a_1, b_2) or (b_1, a_2), which translates to

$$(x^T_{a_1}(m_{a_1}), x^T_{b_1}(m_{b_1}), x^T_{a_2}(m_{a_2}), x^T_{b_2}(m_{b_2})) = (1,0,0,1) \text{ and } (x^T_{a_1}(m_{a_1}), x^T_{b_1}(m_{b_1}), x^T_{a_2}(m_{a_2}), x^T_{b_2}(m_{b_2})) = (0,1,1,0)$$

respectively. Convergence to Nash is expected from the literature since Stochastic Fictitious Play (which is baseline Hedge with $\alpha = 0, \beta < \infty$) shows convergence to Nash regardless of the initial conditions for this kind of game Hofbauer and Sandholm (2002). Convergence to correlated equilibrium is novel in this context.

Using Proposition 1, it is clear to see that if the system is close enough to one of the pure-strategy correlated equilibria (of which Pure Nash equilibria are a subset of), then it would be attracted towards this equilibrium. We would like to study more basins of attraction to delimit in which areas are certain equilibria more attracting.

Theoretical studies on stochastic fictitious play, which is equivalent to the (uncorrelated) hedge with $\alpha = 0, \beta < +\infty$, have shown convergence to a Nash equilibrium in 2×2 games, aligning with the results on the right Pangallo et al. (2022); Hofbauer and Sandholm (2002).

Result 1. *The message distribution η induces the obedient correlated equilibrium around 10% of the time when $\alpha = 0$. The remaining 90% of the times, it converges to pure NE.*

Now, around $\alpha \approx 1$, we have convergence to obedience around 7% of the time but to correlated equilibrium around 50% of the time. A heuristic interpretation can be given comparing with the Best Response Dynamics ($\alpha = 1, \beta = \infty$) (Matsui, 1992). In this dynamics, players the best response (pure strategy) against the opponent's last action. Analogously, if the best response given the message happens to constitute a correlated equilibrium, and β is large enough so that it surpasses the 99.5% threshold, then it would be seen in our graph correspondingly.

Result 2. *The message distribution η induces the obedient correlated equilibrium around 7% of the time when $\alpha \approx 1$.*

The image also clearly display a staircase-like division in the parameter space between the combination of parameters (α, β) that induce convergence to a correlated equilibrium and those that do not. The ratio α/β seems to be the cause of this phenomenon and is implied by Corollary 1.

Result 3. *Convergence to correlated equilibrium is related to the ratio α/β.*

3.1 Price of Learning

On the other hand, let us compare the previous results with Figure 2. This figure depicts the proportion out of 100 simulations for each parameter combination (α, β) of to either *obedient correlated equilibrium* on the left-hand side, and to one of the correlated equilibrium (obedience or pure Nash) on the right-hand side when using the *theoretical welfare optimal* information distribution η' as the correlation device.

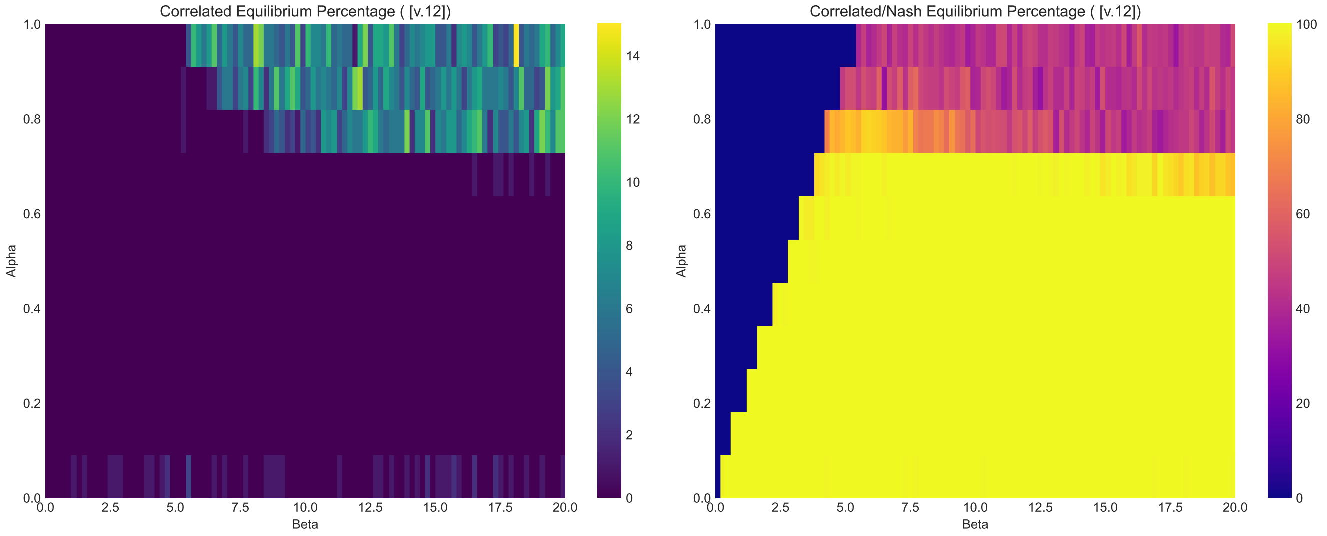


Figure 2: Heatmap of The Correlated Equilibrium Percentage (Excluding Nash Equilibrium) (Left) and Correlated Equilibrium Percentage (Right)

One of the most notable things about this image is almost null convergence to obedience when $\alpha = 0$, in sharp contrast to the previous result induced by η. Nevertheless, it almost always converges to one of the pure Nash equilibria.

This result is reasonable. In this framework, everything is learnt: payoffs, opponent's behavior, correlation between message and opponent's behavior. The theoretically optimal information distribution η' makes the decision-maker indifferent between obeying and deviating, given a message. If, due to randomness, or numerical precision, or to insufficient correlation between the messages and opponent's behavior, the decision-maker becomes more prone to deviating and gets closer to one of the pure Nash, then it becomes attracted and hence, it converges to pure Nash. **?**.

Result 4. *Theoretically optimal information design may not induce obedience.*

Now, comparing the Result 4 to Result 1, we argue that η is "better" than η'. The induced social welfare of the information distributions are:

$$SW_\eta = 9.3, SW_{\eta'} = 9 \tag{9}$$

This paves the way for one idea of Robust Information Design[1]. Instead of employing information design in the usual way, implementing having in mind some uncertainty (stochasticity, inadequate correlation between opponent's action and message). One proposal is to implement the distribution η^* that maximizes:

$$\eta^* \in argmax_{\eta \in \Delta(M)} SW_\eta. \tag{10}$$

Our next challenge is to find an analytical way to calculate η^*. Equivalently, we could define the **Price of Learning (PoL)**, inspired by the Price of Anarchy, Price of Robustness and Price of Stability. Price of Learning would be defined as

$$PoL_\eta = \frac{\text{Theoretically Optimal Welfare - Welfare Induced by } \eta}{\text{Theoretically Optimal Welfare}}. \tag{11}$$

Now, it is interesting to see that $PoL_\eta = \frac{1.2}{10.5}$, part of it induced by how stable and attracting are the Pure Nash equilibria. So, one way to induce a higher PoL is to start $Q^0_{a_i}(m_i)$ not at 0 but close to the desired equilibrium. The interpretation is that the initial values are not zero if the decision-maker has some information about the problem, or the relation between opponent's correlation devices and actions Camerer and Hua Ho (1999).

Experimental evidence, for example, testifies to the effectiveness of public signals instead of private (Ziegler, 2023; Bone et al., 2024; Friedman et al., 2022), of explicit correlation devices (Duffy et al., 2017). There is evidence that belief about correlation is also relevant to achieve correlated equilibrium (Cason and Sharma, 2007; Cason et al., 2020). All of these shift change the starting value $Q^0_{a_i}(m_i)$. One could interpret it as that the decision-maker has mentally played before the game starts.

4 Extensions

Now, this section includes rough ideas that point to possible extensions of our framework.

Proposition 3. *Let Γ be a 2×2 game and consider the Correlated Hedge. For player $i \in 1, 2$ with action space $A_i = \{a_i, b_i\}$, the choice probability $x^t_{a_i}(m_i)$ at period t depends solely on the attraction differential $\Delta Q^t_i(m_i) := Q^t_{a_i}(m_i) - Q^t_{b_i}(m_i)$.*

Proof. Let $\gamma \in [0, 1]$ and there are only 2 actions: $\{a, b\}$ and let $\beta Q^t_{a_i}(m_i) = A, \beta Q^t_{b_i}(m_i) = B$:

$$x^t_{a_i}(m_i) \equiv \frac{exp(\beta Q^t_{a_i}(m_i))}{\sum_{a'_i \in A_i} exp(\beta Q^t_{a'_i}(m_i))} = \gamma$$

$$\gamma = \frac{e^A}{e^A + e^B}$$

$$\gamma(e^A + e^B) = e^A$$

$$\gamma e^B = (1 - \gamma) e^A$$

$$\frac{e^B}{e^A} = \frac{(1 - \gamma)}{\gamma}$$

[1] The terms is already used by Feng et al. (2024) in a slightly different context but similar in spirit

$$B - A = \ln\left(\frac{(1-\gamma)}{\gamma}\right)$$

$$\beta(Q_i^b(t) - Q_i^a(t)) = \ln\left(\frac{(1-\gamma)}{\gamma}\right)$$

$$\beta(Q_i^a(t) - Q_i^b(t)) = \ln\left(\frac{\gamma}{1-\gamma}\right)$$

□

Corollary 1. *In order to have a pure strategy given message, one needs either $\beta = +\infty$ and/or infinite difference of attractions $|Q_i^a(t) - Q_i^b(t)|$, which is only possible if $\alpha = 0$.*

Proof. **Sum of rewards formulation:**

Let $\{t_k\}$ be the sequence of times when $m_i^{t_k} = m_i$, then can rewrite the $Q-$value as the sum of rewards

$$Q_{a_i}^{(n)}(m_i) = \sum_{k=1}^{n}(1-\alpha)^{(n-k)}u_i(a_i, a_{-i}^{t_k}) \tag{12}$$

where $n = \tau_{m_i}(t)$ is the number of times message m_i has appeared up to time t.

Now, let u_{a_i} represent $\max_{a_{-i}\in A_{-i}} u_i(a_i, a_{-i})$ and u_{b_i} represent $\min_{a_{-i}\in A_{-i}} u_i(b_i, a_{-i})$. Then

$$Q_{a_i}^{(n)}(m_i) - Q_{b_i}^{(n)}(m_i) \leq \sum_{k=1}^{n}(1-\alpha)^{(n-k)}[u_{a_i} - u_{b_i}] \leq \lim_{k\to\infty}\sum_{k=1}^{n}(1-\alpha)^{(n-k)}[u_{a_i} - u_{b_i}] = \frac{[u_{a_i} - u_{b_i}]}{\alpha},$$

which is finite if $\alpha > 0$.

□

Now, this points to the idea that **obedience** may only occur strictly in extreme parametrizations of correlated hedge. Descriptively, it is unreasonable to think that decision-makers have full memory, $\alpha = 0$, or that they do not employ any exploration, $\beta = +\infty$.

One natural extension is the use of **adaptive algorithms**. We consider two types of extensions. First, algorithms that reduces exploration over time, so β_t is increasing over time. Second, algorithms in which the information distribution η_t changes over time.

In the bandit literature, it is shown that exploration with fixed rates β is suboptimal when the environment samples actions from fixed probability distributions Cesa-Bianchi et al. (2017). The intuition is that given enough exploration and understanding of the random variables, it is optimal to exploit more than explore. The assumption of fixed environment is not true in our game-theoretic context, since actions of the opponents do change over time. However, it suggests that a dynamically adaptive exploration scheme is worth investigating.

Our first attempt of an algorithm with decreasing exploration in Figure 3. This variation linearly increases β, according to $\beta_t = \beta_0 + kt$, akin to the approach used in Calvano et al. (2020). The aim is to allow agents to explore extensively before converging to a pure action. As seen in the left heatmap, this approach always converges to a pure-action profile given messages. Unfortunately, here there are no evident patterns explaining these results, but they are consistent with the result of the convergence of stochastic fictitious play into a Nash equilibrium.

4.1 Adaptive Designer Theory

This subsection is dedicated to defining a theory in which the information designer also uses a learning algorithm to adapt the correlation device over time to maximize (myopically) the social welfare. The underlying idea is to treat the designer as another player of the game. The motivation behind this idea is the possible ignorance of the information designer of the primitives of the game. *We still do not have clear results on this framework.*

We consider two primary measures for evaluating outcome desirability: social welfare and fairness. Social welfare is defined as

$$SW = \sum_i \sum_m \sum_a \eta(m) x_{a_i}^t(m_i) u_i(a_i, x_{a_{-i}}^t),$$

while fairness as

$$F = \min_i \sum_m \sum_a \eta(m) x_{a_i}^t(m_i) u_i(a, x_{a_{-i}}^t).$$

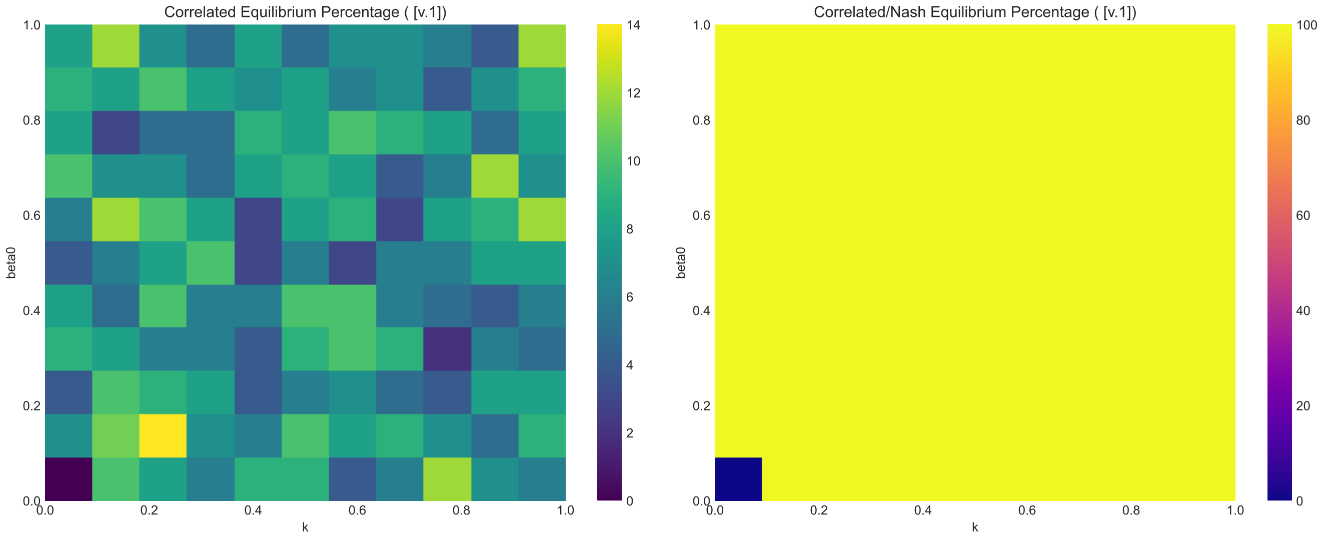


Figure 3: Decreasing Exploration Algorithm: Heatmap of The Correlated Equilibrium Percentage (Excluding Nash Equilibrium) (Left) and Correlated Equilibrium Percentage (Right)

In the generic case, we adapt the message distribution over time: we fine-tune the probability distribution of the messages over time to nudge the learning dynamics towards desirable outcomes. To accomplish this, we introduce a measure of efficiency of a mixed strategy, defined as a map $k : \times_i \Delta(A_i) \to \mathbb{R}$. A representative example of such a mapping is the social welfare induced by a mixed strategy profile, $k(x_1, ..., x_n) = \sum_i \sum_a x(a) u_i(a)$. At iteration t, each message $m \in M$ induces a mixed strategy profile $x^t(m)$. The welfare associated with message $m \in M$ is thus calculated as $k^t(m) = k(x_1^t(m), ..., x_n^t(m))$. To fine-tune the message distribution over time, we implement the following update procedure. First, we update the welfare estimate for each message as follows,

$$W^t(m) = (1 - \alpha) W^{t-1}(m) + k^t(m).$$

Then, we derive the probability of drawing message m at iteration t, denoted as $\eta^t(m)$, through

$$\eta^t(m) = \frac{\exp\left(\beta W^t(m)\right)}{\sum_m \exp\left(\beta W^t(m)\right)}.$$

This approach operates under the assumption that mixed strategies are observable by the algorithm designer. An alternative procedure would involve utilizing only the actions that were actually played. This update rule enables systematic adjustment of message probabilities, shifting weight toward messages that demonstrate superior performance according to the chosen efficiency criterion.

5 Stochastic Approximation

We are currently working on applying stochastic approximation theory to have convergence guarantees on the learning algorithms. Unfortunately, the classical theory of Robbins-Monro (Robbins and Monro, 1951; Benaïm and Weibull, 2003) does not apply to our algorithm. Therefore, we plan to study the current techniques of Falniowski and Mertikopoulos (2025) to understand convergence, using what they call second-order effects.

On the other hand, we have studied which correlation-augmented algorithms that fall into the analytical framework classical theory of stochastic approximation theory. Let us define the **Correlated Discounted-sum** algorithm, proposed by Nicolas:

$$Q_{a_i}^t(m) = (1 - \alpha_t) Q_{a_i}^{t-1}(m) + \alpha_t u_i(a_i, a_{-i}^t)$$

$$x_{a_i}^t(m_i) = \mathbb{1}_{\{m_i^t = m_i\}} \frac{\exp(\beta Q_{a_i}^t(m_i))}{\sum_{a_i' \in A_i} \exp(\beta Q_{a_i'}^t(m_i))}$$

where α_t decreases over time, $\alpha_t \propto \frac{1}{t^\phi}, \phi > 0$. Our analysis have concluded that we can apply the results from stochastic approximation to this dynamics.

Furthermore, let us define Correlated Experience-Weighted Attraction (CEWA), modification of the algorithm proposed by Camerer and Hua Ho (1999). This is a very general learning algorithm, in which Hedge is a subset thereof. The difference is the incorporation of a state variable called Experience $N^t(m_i)$ and the parameters $\rho \in [0,1]$, which control the experience, and $\delta \in [0,1]$ which describes how much we update for counterfactual utilities (the idea is: given a message, do I update the attraction towards an action with the utility that I would have obtained given the opponent's action at time t?). In our case of Correlated Hedge, we take $\delta = 1$ (called full-feedback).

$$\mathcal{N}^t(m_i) = \mathcal{N}^{t-1}(m_i) + \mathbb{1}_{\{m_i^t = m_i\}}(\rho - 1)[\mathcal{N}^{t-1}(m_i) + 1] \tag{Experience}$$

$$\begin{aligned} Q^t_{a_i}(m_i) = Q^{t-1}_{a_i}(m_i) + \frac{\mathbb{1}_{\{m_i^t = m_i\}}}{\mathcal{N}^t(m_i)} \Bigg[& \left((1-\alpha)\mathcal{N}^{t-1}(m_i) - \mathcal{N}^t(m_i)\right) Q^t_{a_i}(m_i) \\ & + \left[\delta + (1-\delta)\mathbb{1}_{\{a_i^t = a_i\}}\right] u_i(a_i, a^t_{-i}) \Bigg] \end{aligned} \tag{Attraction}$$

$$x^t_{a_i}(m_i) = \mathbb{1}_{\{m_i^t = m_i\}} \frac{\exp\left(\beta Q^t_{a_i}(m_i)\right)}{\sum_{a_i' \in A_i} \exp\left(\beta Q^t_{a_i'}(m_i)\right)} \tag{Softmax}$$

The analysis indicates that the stochastic approximation works only if $\rho = 1$. On the other hand, Correlated Hedge uses $\rho = 0$.

6 Bayes-Nash Equilibrium

We are also currently working on the suggestion given by Tristan and Frédéric on analyzing the extended game in the context of a Bayesian game. In this framework, strategies $\sigma_i : M_i \to \Delta(A_i)$ are a full mapping from messages to (potentially mixed) actions. The agent has a probability $x^t_{\sigma_i}$ of choosing a certain strategy. Then a message $m_i \in \eta$ is sampled, so the action $\sigma_i(m_i)$ is chosen, which affects $x^{t+1}_{\sigma_i}$. Further effort is needed to understand more the implications of the model and to define it more properly.

7 Conclusion

We introduce an extension of independent reinforcement learning algorithms in which interactions are mediated by a correlation device. The algorithm we propose enables the emergence of new outcomes – specifically, new correlated strategies – through the pushforward of a message distribution and the players' conditional mixed strategies. We consider both fixed and adaptive message distributions, with the latter allowing for outcome refinement via adjustments based on observed play and a welfare criterion. This mechanism nudges the selection of desirable outcomes and improves upon standard independent learning.

Several open questions remain. For instance, under what conditions on payoffs and message structures does the procedure guarantee to outperform baseline algorithms in selecting more desirable outcomes? How do convergence properties depend on the structure of the payoff matrix? From a theoretical perspective, we are exploring whether the resulting outcomes correspond to Nash equilibria of Bayesian games extended with some information structures. On the empirical side, we are investigating the robustness of simulations, particularly how increasing the number of messages affects convergence and equilibrium selection.

Finally, identifying meaningful economic applications is a priority for future work, with a focus on validating the model in environments where coordination and correlated behavior are central.

References

Simon P Anderson, Andre De Palma, and Jacques-Francois Thisse. *Discrete choice theory of product differentiation.* MIT press, 1992.

Robert J Aumann. Correlated equilibrium as an expression of bayesian rationality. *Econometrica: Journal of the Econometric Society*, pages 1–18, 1987.

James P Bailey and Georgios Piliouras. Multiplicative weights update in zero-sum games. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 321–338, 2018.

Siddharth Barman and Katrina Ligett. Finding any nontrivial coarse correlated equilibrium is hard. *ACM SIGecom Exchanges*, 14(1):76–79, 2015.

Moshe E Ben-Akiva and Steven R Lerman. *Discrete choice analysis: theory and application to travel demand*, volume 9. MIT press, 1985.

Michel Benaïm and Jörgen W Weibull. Deterministic approximation of stochastic evolution in games. *Econometrica*, 71(3):873–903, 2003.

John Bone, Michalis Drouvelis, Zeynep Gürgüç, and Indrajit Ray. Following recommendations from public and private correlation devices in a game of chicken. Technical report, Cardiff Economics Working Papers, 2024.

Emilio Calvano, Giacomo Calzolari, Vincenzo Denicolo, and Sergio Pastorello. Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review*, 110(10):3267–3297, 2020.

Colin Camerer and Teck Hua Ho. Experience-weighted attraction learning in normal form games. *Econometrica*, 67 (4):827–874, 1999.

Ilayda Canyakmaz, Iosif Sakos, Wayne Lin, Antonios Varvitsiotis, and Georgios Piliouras. Learning and steering game dynamics towards desirable outcomes. *arXiv preprint arXiv:2404.01066*, 2024.

Timothy N Cason and Tridib Sharma. Recommended play and correlated equilibria: an experimental study. *Economic Theory*, 33(1):11–27, 2007.

Timothy N Cason, Tridib Sharma, and Radovan Vadovič. Correlated beliefs: Predicting outcomes in 2× 2 games. *Games and Economic Behavior*, 122:256–276, 2020.

David Cass and Karl Shell. Do sunspots matter? *Journal of political economy*, 91(2):193–227, 1983.

Jakub Černỳ, Bo An, and Allan N Zhang. Quantal correlated equilibrium in normal form games. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 210–239, 2022.

Nicolò Cesa-Bianchi, Claudio Gentile, Gábor Lugosi, and Gergely Neu. Boltzmann exploration done right. *Advances in neural information processing systems*, 30, 2017.

Johanne Cohen, Amélie Héliou, and Panayotis Mertikopoulos. Hedging under uncertainty: Regret minimization meets exponentially fast convergence. In *International Symposium on Algorithmic Game Theory*, pages 252–263. Springer, 2017.

Arthur Dolgopolov. Reinforcement learning in a prisoner's dilemma. *Games and Economic Behavior*, 144:84–103, 2024.

John Duffy and Eric O'N Fisher. Sunspots in the laboratory. *American Economic Review*, 95(3):510–529, 2005.

John Duffy, Ernest K Lai, and Wooyoung Lim. Coordination via correlation: An experimental study. *Economic Theory*, 64:265–304, 2017.

Federico Echenique and Kota Saito. General luce model. *Economic Theory*, 68(4):811–826, 2019.

Ido Erev and Alvin E Roth. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American economic review*, pages 848–881, 1998.

Fryderyk Falniowski and Panayotis Mertikopoulos. On the discrete-time origins of the replicator dynamics: From convergence to instability and chaos. *International Journal of Game Theory*, 54(1):7, 2025.

Yiding Feng, Chien-Ju Ho, and Wei Tang. Rationality-robust information design: Bayesian persuasion under quantal response. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 501–546. SIAM, 2024.

Françoise Forges. Communication equilibria in repeated games with incomplete information. *Mathematics of Operations Research*, 13(2):191–231, 1988.

Dean P Foster and Rakesh V Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21(1-2):40–55, 1997.

Daniel Friedman, Jean Paul Rabanal, Olga A Rud, and Shuchen Zhao. On the empirical relevance of correlated equilibrium. *Journal of Economic Theory*, 205:105531, 2022.

Drew Fudenberg and David K Levine. *The theory of learning in games*, volume 2. MIT press, 1998.

Tobias Galla and J. Doyne Farmer. Complex dynamics in learning complicated games. *Proceedings of the National Academy of Sciences of the United States of America*, 110(4):1232–1236, 1 2013. ISSN 00278424. doi: 10.1073/pnas.1109672110.

Jacob K Goeree and Charles A Holt. Ten little treasures of game theory and ten intuitive contradictions. *American Economic Review*, 91(5):1402–1422, 2001.

Amy Greenwald, Keith Hall, Roberto Serrano, et al. Correlated q-learning. In *ICML*, volume 3, pages 242–249, 2003.

Sergiu Hart and Andreu Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5):1127–1150, 2000.

Morris W Hirsch, Stephen Smale, and Robert L Devaney. *Differential equations, dynamical systems, and an introduction to chaos*. Academic press, 2013.

Josef Hofbauer and William H Sandholm. On the global convergence of stochastic fictitious play. *Econometrica*, 70 (6):2265–2294, 2002.

Josef Hofbauer and Karl Sigmund. Evolutionary game dynamics. *Bulletin of the American mathematical society*, 40 (4):479–519, 2003.

Stefanos Leonardos and Georgios Piliouras. Exploration-exploitation in multi-agent learning: Catastrophe theory meets game theory. *Artificial Intelligence*, 304:103653, 2022.

R Duncan Luce et al. *Individual choice behavior*, volume 4. Wiley New York, 1959.

Jason R Marden. Selecting efficient correlated equilibria through distributed learning. *Games and Economic Behavior*, 106:114–133, 2017.

Akihiko Matsui. Best response dynamics and socially stable strategies. *Journal of Economic Theory*, 57(2):343–362, 1992.

Roger B Myerson. Multistage games with communication. *Econometrica: Journal of the Econometric Society*, pages 323–358, 1986.

Marco Pangallo, James BT Sanders, Tobias Galla, and J Doyne Farmer. Towards a taxonomy of learning dynamics in 2× 2 games. *Games and Economic Behavior*, 132:1–21, 2022.

Herbert Robbins and Sutton Monro. A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407, 1951.

Yuzuru Sato, Eizo Akiyama, and J. Doyne Farmer. Chaos in learning a simple two-person game. *Proceedings of the National Academy of Sciences of the United States of America*, 99(7):4748–4751, 4 2002. ISSN 00278424. doi: 10.1073/pnas.032086299.

Peter Schuster and Karl Sigmund. Replicator dynamics. *Journal of theoretical biology*, 100(3):533–538, 1983.

Yoav Shoham, Rob Powers, and Trond Grenager. If multi-agent learning is the answer, what is the question? *Artificial intelligence*, 171(7):365–377, 2007.

Brian Hu Zhang, Gabriele Farina, Ioannis Anagnostides, Federico Cacciamani, Stephen Marcus McAleer, Andreas Alexander Haupt, Andrea Celli, Nicola Gatti, Vincent Conitzer, and Tuomas Sandholm. Steering no-regret learners to a desired equilibrium. *arXiv preprint arXiv:2306.05221*, 2023.

Andreas Ziegler. Persuading an audience: Testing information design in the laboratory. Technical report, Tinbergen Institute Discussion Paper, 2023.

8 Appendix

9 General Case Jacobian

We want to define the Jacobian, so it is useful to calculate:

$$\frac{\partial u_i^t(a_i|m_i)}{\partial x_{a_{-i}}^t(m_{-i})} = \eta(m_{-i}|m_i)[u_i(a_i,a_{-i}) - u_i(a_i,a'_{-i}))]$$

$$\frac{\partial u_i^t(a_i|m_i)}{\partial x_{a'_{-i}}^t(m'_{-i})} = \eta(m'_{-i}|m_i)[u_i(a_i,a'_{-i}) - u_i(a_i,a_{-i}))]$$

$$\frac{\partial u_i^t(a'_i|m_i)}{\partial x_{a_{-i}}^t(m_{-i})} = \eta(m_{-i}|m_i)[u_i(a'_i,a_{-i}) - u_i(a'_i,a'_{-i}))]$$

$$\frac{\partial u_i^t(a'_i|m_i)}{\partial x_{a'_{-i}}^t(m'_{-i})} = \eta(m'_{-i}|m_i)[u_i(a'_i,a_{-i}) - u_i(a'_i,a'_{-i}))]$$

Now, we can write the partials

$$\frac{\partial \dot{x}_{a_i}^t(m_i)}{\partial x_{a_i}^t(m_i)} = \beta\eta(m_i)(1-2x_{a_i}^t(m_i))[u_i^t(a_i|m_i) - u_i^t(a'_i|m_i)] + \alpha\eta(m_i)(1-2x_{a_i}^t(m_i))\ln\left(\frac{x_{a_i}^t(m_i)}{1-x_{a_i}^t(m_i)}\right) + \alpha\eta(m_i)$$

$$\frac{\partial \dot{x}_{a_i}^t(m_i)}{\partial x_{a_{-i}}^t(m_{-i})} = \beta x_{a_i}^t(m_i)\eta(m_i)(1-x_{a_i}^t(m_i))\eta(m_{-i}|m_i)\left[u_i(a_i,a_{-i}) - u_i(a_i,a'_{-i}) - u_i(a'_i,a_{-i}) + u_i(a'_i,a'_{-i})\right]$$

$$\frac{\partial \dot{x}_{a_i}^t(m_i)}{\partial x_{a'_{-i}}^t(m'_{-i})} = \beta x_{a_i}^t(m_i)\eta(m_i)(1-x_{a_i}^t(m_i))\eta(m'_{-i}|m_i)\left[u_i(a'_i,a_{-i}) - u_i(a'_i,a'_{-i}) - u_i(a_i,a_{-i}) + u_i(a_i,a'_{-i})\right]$$

Now, for the other partials, we have:

$$\frac{\partial \dot{x}_{a'_i}^t(m'_i)}{\partial x_{a'_i}^t(m'_i)} = \beta\eta(m'_i)(1-2x_{a'_i}^t(m'_i))[u_i^t(a'_i|m'_i) - u_i^t(a_i|m'_i)] + \alpha\eta(m'_i)(1-2x_{a'_i}^t(m'_i))\ln\left(\frac{x_{a'_i}^t(m'_i)}{1-x_{a'_i}^t(m'_i)}\right) + \alpha\eta(m'_i)$$

$$\frac{\partial \dot{x}_{a'_i}^t(m'_i)}{\partial x_{a_{-i}}^t(m_{-i})} = \beta x_{a'_i}^t(m'_i)\eta(m'_i)(1-x_{a'_i}^t(m'_i))\eta(m_{-i}|m'_i)\left[u_i(a'_i,a_{-i}) - u_i(a'_i,a'_{-i}) - u_i(a_i,a_{-i}) + u_i(a_i,a'_{-i})\right]$$

$$\frac{\partial \dot{x}_{a'_i}^t(m'_i)}{\partial x_{a'_{-i}}^t(m'_{-i})} = \beta x_{a'_i}^t(m'_i)\eta(m'_i)(1-x_{a'_i}^t(m'_i))\eta(m'_{-i}|m'_i)\left[u_i(a_i,a_{-i}) - u_i(a_i,a'_{-i}) - u_i(a'_i,a_{-i}) + u_i(a'_i,a'_{-i})\right]$$

We also have

$$\frac{\partial \dot{x}_{a_{-i}}^t(m_{-i})}{\partial x_{a_{-i}}^t(m_{-i})} = \beta\eta(m_{-i})(1-2x_{a_{-i}}^t(m_{-i}))[u_{-i}^t(a_{-i}|m_{-i}) - u_{-i}^t(a'_{-i}|m_{-i})] + \alpha\eta(m_{-i})(1-2x_{a_{-i}}^t(m_{-i}))\ln\left(\frac{x_{a_{-i}}^t(m_{-i})}{1-x_{a_{-i}}^t(m_{-i})}\right) + \alpha\eta(m_{-i}$$

$$\frac{\partial \dot{x}_{a_{-i}}^t(m_{-i})}{\partial x_{a_i}^t(m_i)} = \beta x_{a_{-i}}^t(m_{-i})\eta(m_{-i})(1-x_{a_{-i}}^t(m_{-i}))\eta(m_i|m_{-i})\left[u_{-i}(a_{-i},a_i) - u_{-i}(a_{-i},a'_i) - u_i(a'_{-i},a_i) + u_i(a'_{-i},a'_i)\right]$$

$$\frac{\partial \dot{x}_{a_{-i}}^t(m_{-i})}{\partial x_{a'_i}^t(m'_i)} = \beta x_{a_{-i}}^t(m_{-i})\eta(m_{-i})(1-x_{a_{-i}}^t(m_{-i}))\eta(m'_i|m_{-i})\left[u_{-i}(a'_{-i},a_i) - u_{-i}(a'_{-i},a'_i) - u_{-i}(a_{-i},a_i) + u_{-i}(a_{-i},a'_i)\right]$$

And finally,

$$\frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x^t_{a'_{-i}}(m'_{-i})} = \beta\eta(m'_{-i})(1-2x^t_{a'_{-i}}(m'_{-i}))[u^t_{-i}(a'_{-i}|m'_{-i})-u^t_{-i}(a_{-i}|m'_{-i})]+\alpha\eta(m'_{-i})(1-2x^t_{a'_{-i}}(m'_{-i}))\ln\left(\frac{x^t_{a'_{-i}}(m'_{-i})}{1-x^t_{a'_{-i}}(m'_{-i})}\right)+\alpha\eta(m'_{-i}$$

$$\frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x^t_{a_i}(m_i)} = \beta x^t_{a'_{-i}}(m'_{-i})\eta(m'_{-i})(1-x^t_{a'_{-i}}(m'_{-i}))\eta(m_i|m'_{-i})\left[u_{-i}(a'_{-i},a_i)-u_{-i}(a'_{-i},a'_i)-u_i(a_{-i},a_i)+u_i(a_{-i},a'_i)\right]$$

$$\frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x^t_{a'_i}(m'_i)} = \beta x^t_{a'_{-i}}(m'_{-i})\eta(m'_{-i})(1-x^t_{a_{-i}}(m'_{-i}))\eta(m'_i|m'_{-i})\left[u_{-i}(a_{-i},a_i)-u_{-i}(a_{-i},a'_i)-u_{-i}(a'_{-i},a_i)+u_{-i}(a'_{-i},a'_i)\right]$$

Proof. Suppose $x^t_{a_i}(m_i)=0$, then

$$\dot{x}^t_{a_i}(m_i) = 0\ln(0).$$

Suppose $x^t_{a_i}(m_i)=1$, then

$$\dot{x}^t_{a_i}(m_i) = \beta\eta(m_i)\left[u^t_i(a_i \mid m_i) - u^t_i(a_i \mid m_i)\right] - \alpha\eta(m_i)\left[\ln(1)-\ln(1)\right] = 0.$$

Concerning stability, one way to prove stability is through **Lyapunov Linearization Theorem** which states that if all eigenvalues of the Jacobian have strictily negative parts, then it is asymptotically stable Hirsch et al. (2013).

And since we are in 2 × 2 game, we could write the continuous-time equivalent as

$$\begin{aligned}\dot{x}^t_{a_i}(m_i) = \beta x^t_{a_i}(m_i)\eta(m_i)(1-x^t_{a_i}(m_i))\left[u^t_i(a_i|m_i)-u^t_i(a'_i|m_i)\right] - \\ \alpha x^t_{a_i}(m_i)\eta(m_i)(1-x^t_{a_i}(m_i))\left[\ln\left(\frac{x^t_{a_i}(m_i)}{1-x^t_{a_i}(m_i)}\right)\right]\end{aligned}$$

and the average utility of playing a_i given m_i as

$$\begin{aligned}u^t_i(a_i|m_i) = \eta(m_{-i}|m_i)[u_i(a_i,a_{-i})x^t_{a_{-i}}(m_{-i})+u_i(a_i,a'_{-i})(1-x^t_{a_{-i}}(m_{-i}))]+ \\ \eta(m'_{-i}|m_i)[u_i(a_i,a_{-i})(1-x^t_{a'_{-i}}(m'_{-i}))+u_i(a_i,a'_{-i})x_{a'_{-i}}(m'_{-i}))].\end{aligned}$$

We want to define the Jacobian, so it is useful to calculate:

$$\begin{aligned}\frac{\partial u^t_i(a_i|m_i)}{\partial x^t_{a_{-i}}(m_{-i})} &= \eta(m_{-i}|m_i)[u_i(a_i,a_{-i})-u_i(a_i,a'_{-i}))] \\ \frac{\partial u^t_i(a_i|m_i)}{\partial x^t_{a'_{-i}}(m'_{-i})} &= \eta(m'_{-i}|m_i)[u_i(a_i,a'_{-i})-u_i(a_i,a_{-i}))] \\ \frac{\partial u^t_i(a'_i|m_i)}{\partial x^t_{a_{-i}}(m_{-i})} &= \eta(m_{-i}|m_i)[u_i(a'_i,a_{-i})-u_i(a'_i,a'_{-i}))] \\ \frac{\partial u^t_i(a'_i|m_i)}{\partial x^t_{a'_{-i}}(m'_{-i})} &= \eta(m'_{-i}|m_i)[u_i(a'_i,a_{-i})-u_i(a'_i,a'_{-i}))]\end{aligned}$$

Now, we can write the partials

$$\frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x^t_{a_i}(m_i)} = \beta\eta(m_i)(1-2x^t_{a_i}(m_i))[u^t_i(a_i|m_i)-u^t_i(a'_i|m_i)]+\alpha\eta(m_i)(1-2x^t_{a_i}(m_i))\ln\left(\frac{x^t_{a_i}(m_i)}{1-x^t_{a_i}(m_i)}\right)+\alpha\eta(m_i)$$

$$\frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x^t_{a_{-i}}(m_{-i})} = \beta x^t_{a_i}(m_i)\eta(m_i)(1 - x^t_{a_i}(m_i))\eta(m_{-i}|m_i)\left[u_i(a_i, a_{-i}) - u_i(a_i, a'_{-i}) - u_i(a'_i, a_{-i}) + u_i(a'_i, a'_{-i})\right]$$

$$\frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x^t_{a'_{-i}}(m'_{-i})} = \beta x^t_{a_i}(m_i)\eta(m_i)(1 - x^t_{a_i}(m_i))\eta(m'_{-i}|m_i)\left[u_i(a'_i, a_{-i}) - u_i(a'_i, a'_{-i}) - u_i(a_i, a_{-i}) + u_i(a_i, a'_{-i})\right]$$

Now, for the other partials, we have:

$$\frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x^t_{a'_i}(m'_i)} = \beta\eta(m'_i)(1 - 2x^t_{a'_i}(m'_i))[u^t_i(a'_i|m'_i) - u^t_i(a_i|m'_i)] + \alpha\eta(m'_i)(1 - 2x^t_{a'_i}(m'_i))\ln\left(\frac{x^t_{a'_i}(m'_i)}{1 - x^t_{a'_i}(m'_i)}\right) + \alpha\eta(m'_i)$$

$$\frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x^t_{a_{-i}}(m_{-i})} = \beta x^t_{a'_i}(m'_i)\eta(m'_i)(1 - x^t_{a'_i}(m'_i))\eta(m_{-i}|m'_i)\left[u_i(a'_i, a_{-i}) - u_i(a'_i, a'_{-i}) - u_i(a_i, a_{-i}) + u_i(a_i, a'_{-i})\right]$$

$$\frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x^t_{a'_{-i}}(m'_{-i})} = \beta x^t_{a'_i}(m'_i)\eta(m'_i)(1 - x^t_{a'_i}(m'_i))\eta(m'_{-i}|m'_i)\left[u_i(a_i, a_{-i}) - u_i(a_i, a'_{-i}) - u_i(a'_i, a_{-i}) + u_i(a'_i, a'_{-i})\right]$$

We also have

$$\frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x^t_{a_{-i}}(m_{-i})} = \beta\eta(m_{-i})(1{-}2x^t_{a_{-i}}(m_{-i}))[u^t_{-i}(a_{-i}|m_{-i}){-}u^t_{-i}(a'_{-i}|m_{-i})]{+}\alpha\eta(m_{-i})(1{-}2x^t_{a_{-i}}(m_{-i}))\ln\left(\frac{x^t_{a_{-i}}(m_{-i})}{1 - x^t_{a_{-i}}(m_{-i})}\right){+}\alpha\eta(m_{-i}$$

$$\frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x^t_{a_i}(m_i)} = \beta x^t_{a_{-i}}(m_{-i})\eta(m_{-i})(1{-}x^t_{a_{-i}}(m_{-i}))\eta(m_i|m_{-i})\left[u_{-i}(a_{-i}, a_i) - u_{-i}(a_{-i}, a'_i) - u_i(a'_{-i}, a_i) + u_i(a'_{-i}, a'_i)\right]$$

$$\frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x^t_{a'_i}(m'_i)} = \beta x^t_{a_{-i}}(m_{-i})\eta(m_{-i})(1{-}x^t_{a_{-i}}(m_{-i}))\eta(m'_i|m_{-i})\left[u_{-i}(a'_{-i}, a_i) - u_{-i}(a'_{-i}, a'_i) - u_{-i}(a_{-i}, a_i) + u_{-i}(a_{-i}, a'_i)\right]$$

And finally,

$$\frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x^t_{a'_{-i}}(m'_{-i})} = \beta\eta(m'_{-i})(1{-}2x^t_{a'_{-i}}(m'_{-i}))[u^t_{-i}(a'_{-i}|m'_{-i}){-}u^t_{-i}(a_{-i}|m'_{-i})]{+}\alpha\eta(m'_{-i})(1{-}2x^t_{a'_{-i}}(m'_{-i}))\ln\left(\frac{x^t_{a'_{-i}}(m'_{-i})}{1 - x^t_{a'_{-i}}(m'_{-i})}\right){+}\alpha\eta(m'_{-i}$$

$$\frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x^t_{a_i}(m_i)} = \beta x^t_{a'_{-i}}(m'_{-i})\eta(m'_{-i})(1{-}x^t_{a'_{-i}}(m'_{-i}))\eta(m_i|m'_{-i})\left[u_{-i}(a'_{-i}, a_i) - u_{-i}(a'_{-i}, a'_i) - u_i(a_{-i}, a_i) + u_i(a_{-i}, a'_i)\right]$$

$$\frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x^t_{a'_i}(m'_i)} = \beta x^t_{a'_{-i}}(m'_{-i})\eta(m'_{-i})(1{-}x^t_{a_{-i}}(m'_{-i}))\eta(m'_i|m'_{-i})\left[u_{-i}(a_{-i}, a_i) - u_{-i}(a_{-i}, a'_i) - u_{-i}(a'_{-i}, a_i) + u_{-i}(a'_{-i}, a'_i)\right]$$

The Jacobian is defined as follows:

$$J(x) = \begin{bmatrix} \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a'_{-i}}(m'_{-i})} \\ \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_{-i}}(m'_{-i})} \\ \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a'_{-i}}(m'_{-i})} \\ \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a'_{-i}}(m'_{-i})} \end{bmatrix}.$$

It follows that:

$$J(x) = \begin{bmatrix} \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_i}(m_i)} & 0 & \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a'_{-i}}(m'_{-i})} \\ 0 & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_{-i}}(m'_{-i})} \\ \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a_{-i}}(m_{-i})} & 0 \\ \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a'_i}(m'_i)} & 0 & \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a'_{-i}}(m'_{-i})} \end{bmatrix}.$$

To compute the eigenvalues, we divide the Jacobian in

$$J(x) = \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

Hence, the eigenvalues are defined by the formula (Schur's formula)

$$\det(J(x) - \lambda I) = \det(A - \lambda I_2) \cdot \det(D - \lambda I_2 - C(A - \lambda I_2)^{-1} B)$$

$$\det(A - \lambda I_2) = \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_i}(m_i)} \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_i}(m'_i)} - \lambda \left(\frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_i}(m_i)} + \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_i}(m'_i)} \right) + \lambda^2$$

Also, we have

$$(A - \lambda I_2)^{-1} = \frac{1}{\left| \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_i}(m_i)} \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_i}(m'_i)} - \lambda \left(\frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_i}(m_i)} + \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_i}(m'_i)} \right) + \lambda^2 \right|} \begin{bmatrix} \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_i}(m'_i)} & 0 \\ 0 & \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_i}(m_i)} \end{bmatrix}$$

[HERE]

$$J(x) = \begin{bmatrix} \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_i}(m_i)} & 0 & \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a'_{-i}}(m'_{-i})} \\ 0 & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_{-i}}(m'_{-i})} \\ \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a_{-i}}(m_{-i})} & 0 \\ \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a'_i}(m'_i)} & 0 & \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a'_{-i}}(m'_{-i})} \end{bmatrix}$$

We write this as a block matrix:

$$J = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where:

$$A = \begin{bmatrix} a & 0 \\ 0 & a' \end{bmatrix}, \quad D = \begin{bmatrix} d & 0 \\ 0 & d' \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}, \quad C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

We compute the characteristic polynomial using the Schur complement:

$$\chi(\lambda) = \det(J - \lambda I_2) = \det(A_\lambda) \cdot \det\left(D_\lambda - C A_\lambda^{-1} B\right)$$

Let:

$$A_\lambda = A - \lambda I = \begin{bmatrix} a - \lambda & 0 \\ 0 & a' - \lambda \end{bmatrix}, \quad D_\lambda = D - \lambda I = \begin{bmatrix} d - \lambda & 0 \\ 0 & d' - \lambda \end{bmatrix}, \quad D_\lambda^{-1} = \begin{bmatrix} \frac{1}{d-\lambda} & 0 \\ 0 & \frac{1}{d'-\lambda} \end{bmatrix}$$

Now, calculating, we have

$$A_\lambda = \begin{bmatrix} a - \lambda & 0 \\ 0 & a' - \lambda \end{bmatrix}, \quad A_\lambda^{-1} = \begin{bmatrix} \frac{1}{a-\lambda} & 0 \\ 0 & \frac{1}{a'-\lambda} \end{bmatrix}$$

$$B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}, \quad C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

First, compute the intermediate product:

$$CA_\lambda^{-1} = \begin{bmatrix} \frac{c_{11}}{a-\lambda} & \frac{c_{12}}{a'-\lambda} \\ \frac{c_{21}}{a-\lambda} & \frac{c_{22}}{a'-\lambda} \end{bmatrix}$$

Now compute the final product:

$$CA_\lambda^{-1}B = \begin{bmatrix} \dfrac{c_{11}b_{11}}{a-\lambda} + \dfrac{c_{12}b_{21}}{a'-\lambda} & \dfrac{c_{11}b_{12}}{a-\lambda} + \dfrac{c_{12}b_{22}}{a'-\lambda} \\ \dfrac{c_{21}b_{11}}{a-\lambda} + \dfrac{c_{22}b_{21}}{a'-\lambda} & \dfrac{c_{21}b_{12}}{a-\lambda} + \dfrac{c_{22}b_{22}}{a'-\lambda} \end{bmatrix}$$

$$\det(CA_\lambda^{-1}B) = \left(\frac{c_{11}b_{11}}{a-\lambda} + \frac{c_{12}b_{21}}{a'-\lambda}\right)\left(\frac{c_{21}b_{12}}{a-\lambda} + \frac{c_{22}b_{22}}{a'-\lambda}\right) - \left(\frac{c_{11}b_{12}}{a-\lambda} + \frac{c_{12}b_{22}}{a'-\lambda}\right)\left(\frac{c_{21}b_{11}}{a-\lambda} + \frac{c_{22}b_{21}}{a'-\lambda}\right)$$

So, our determinants are

$$(a-\lambda)(a'-\lambda) \cdot \left(\left(\frac{c_{11}b_{11}}{a-\lambda} + \frac{c_{12}b_{21}}{a'-\lambda}\right)\left(\frac{c_{21}b_{12}}{a-\lambda} + \frac{c_{22}b_{22}}{a'-\lambda}\right) - \left(\frac{c_{11}b_{12}}{a-\lambda} + \frac{c_{12}b_{22}}{a'-\lambda}\right)\left(\frac{c_{21}b_{11}}{a-\lambda} + \frac{c_{22}b_{21}}{a'-\lambda}\right)\right)$$

Which simplifies to
Let $\alpha = a - \lambda$, $\beta = a' - \lambda$. Then:

$$\alpha\beta \cdot \left[\left(\frac{c_{11}b_{11}}{\alpha} + \frac{c_{12}b_{21}}{\beta}\right)\left(\frac{c_{21}b_{12}}{\alpha} + \frac{c_{22}b_{22}}{\beta}\right) - \left(\frac{c_{11}b_{12}}{\alpha} + \frac{c_{12}b_{22}}{\beta}\right)\left(\frac{c_{21}b_{11}}{\alpha} + \frac{c_{22}b_{21}}{\beta}\right)\right]$$

Expanding, we get:

$$= \beta(c_{11}b_{11}c_{21}b_{12} - c_{11}b_{12}c_{21}b_{11}) + (c_{11}b_{11}c_{22}b_{22} + c_{12}b_{21}c_{21}b_{12} - c_{11}b_{12}c_{22}b_{21} - c_{12}b_{22}c_{21}b_{11}) + \alpha(c_{12}b_{21}c_{22}b_{22} - c_{12}b_{22}c_{22}b_{21})$$

Therefore, the final result is:

$$(a-\lambda)(a'-\lambda) \cdot \det(CA_\lambda^{-1}B) = \beta(c_{11}b_{11}c_{21}b_{12} - c_{11}b_{12}c_{21}b_{11}) + (c_{11}b_{11}c_{22}b_{22} + c_{12}b_{21}c_{21}b_{12} - c_{11}b_{12}c_{22}b_{21} - c_{12}b_{22}c_{21}b_{11}) + \alpha(c_{12}b_{21}c_{22}b$$

Then the Schur complement is:

$$S(\lambda) = A_\lambda - BD_\lambda^{-1}C$$

NOW: if pure-strategy given message, then only the self-interaction term is non-null, so the Jacobian can be re-written as:

$$J(x) = \begin{bmatrix} \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_i}(m_i)} & 0 & 0 & 0 \\ 0 & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_i}(m'_i)} & 0 & 0 \\ 0 & 0 & \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a_{-i}}(m_{-i})} & 0 \\ 0 & 0 & 0 & \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a'_{-i}}(m'_{-i})} \end{bmatrix}$$

Which means they are the eigenvalues. Which means that their real parts (they are purely real numbers) is negative iff

$$u_i^t(a_i|m_i) - u_i^t(a'_i|m_i) \leq 0$$

NOW,

$$CA_\lambda^{-1}B = \begin{bmatrix} \dfrac{c_{11}b_{11}}{a-\lambda} + \dfrac{c_{12}b_{21}}{a'-\lambda} & \dfrac{c_{11}b_{12}}{a-\lambda} + \dfrac{c_{12}b_{22}}{a'-\lambda} \\ \dfrac{c_{21}b_{11}}{a-\lambda} + \dfrac{c_{22}b_{21}}{a'-\lambda} & \dfrac{c_{21}b_{12}}{a-\lambda} + \dfrac{c_{22}b_{22}}{a'-\lambda} \end{bmatrix}$$

$$D_\lambda - CA_\lambda^{-1}B = \begin{bmatrix} (d-\lambda) - \left(\frac{c_{11}b_{11}}{a-\lambda} + \frac{c_{12}b_{21}}{a'-\lambda}\right) & -\left(\frac{c_{11}b_{12}}{a-\lambda} + \frac{c_{12}b_{22}}{a'-\lambda}\right) \\ -\left(\frac{c_{21}b_{11}}{a-\lambda} + \frac{c_{22}b_{21}}{a'-\lambda}\right) & (d'-\lambda) - \left(\frac{c_{21}b_{12}}{a-\lambda} + \frac{c_{22}b_{22}}{a'-\lambda}\right) \end{bmatrix}$$

$$\det(D_\lambda - CA_\lambda^{-1}B) = (d-\lambda)(d'-\lambda) - (d'-\lambda)\left(\frac{c_{11}b_{11}}{a-\lambda} + \frac{c_{12}b_{21}}{a'-\lambda}\right)$$
$$-(d-\lambda)\left(\frac{c_{21}b_{12}}{a-\lambda} + \frac{c_{22}b_{22}}{a'-\lambda}\right) + \frac{((a'-\lambda)c_{11}b_{11} + (a-\lambda)c_{12}b_{12})\,((a'-\lambda)c_{21}b_{12} + (a-\lambda)c_{22}b_{22})}{(a-\lambda)(a'-\lambda)}$$

Therefore,

$$\det(D_\lambda - CA_\lambda^{-1}B) = (d-\lambda)(d'-\lambda) - (d-\lambda)\left(\frac{c_{21}b_{12}}{a-\lambda} + \frac{c_{22}b_{22}}{a'-\lambda}\right) - (d'-\lambda)\left(\frac{c_{11}b_{11}}{a-\lambda} + \frac{c_{12}b_{21}}{a'-\lambda}\right) + \frac{\det(C)\det(B)}{(a-\lambda)(a'-\lambda)}$$

$$\det(A_\lambda)\det(D_\lambda - CA_\lambda^{-1}B) = (a-\lambda)(a'-\lambda)(d-\lambda)(d'-\lambda) - (d'-\lambda)\,((a'-\lambda)c_{11}b_{11} + (a-\lambda)c_{12}b_{21})$$
$$-(d-\lambda)\,((a'-\lambda)c_{21}b_{12} + (a-\lambda)c_{22}b_{22}) + ((a'-\lambda)c_{11}b_{11} + (a-\lambda)c_{12}b_{12})\,((a'-\lambda)c_{21}b_{12} + (a-\lambda)c_{22}b_{22})$$

The final step that needs validation is the evaluation of the Jacobian. I believe that (asymptotic) stability would be equivalent to correlated equilibria when $\alpha = 0$, given similar results when the algorithm has no messages Pangallo et al. (2022). A correlated equilibrium, characterized by $\forall i \in N, \forall a_i, a_i' \in A_i, \forall m_i \in M_i$:

$$\sum_{a_{-i}\in A_{-i}} \sum_{m_{-i}\in M_{-i}} \eta(m_i, m_{-i}) x^t_{a_i}(m_i) x^t_{a_{-i}}(m_{-i})[u_i(a_i, a_{-i}) - u_i(a_i', a_{-i})] \geq 0. \tag{13}$$

This is equivalent to

$$\sum_{a_{-i}\in A_{-i}} \sum_{m_{-i}\in M_{-i}} \eta(m_i|m_{-i})\eta(m_{-i}) x^t_{a_i}(m_i) x^t_{a_{-i}}(m_{-i})[u_i(a_i, a_{-i}) - u_i^t(a_i', a_{-i})] \geq 0$$

$$\eta(m_i) x^t_{a_i}(m_i)[u^t_i(a_i \mid m_i) - u^t_i(a_i' \mid m_i)] \geq 0$$

□

9.1 Other Games

In this section of the appendix, we apply correlated Hedge into different types of games as a test of robustness.

Example 1. Coordination game with full feedback Consider a standard coordination game with the following payoff structure

	b_1	b_2
a_1	$1, 3$	$0, 0$
a_2	$0, 0$	$3, 1$

We examine an implementation of the static algorithm described previously with a set of public messages $M = \{m_1, m_2\}$ with a uniform probability distribution. In the standard algorithm with independent learners, the outcome invariably converges to one of the two pure strategy Nash equilibria. However, when implementing our extended algorithm with messages, we observe the emergence of a novel correlated outcome that randomizes between the two pure strategy Nash equilibria. This finding demonstrates the capacity of our approach to expand the set of outcomes beyond what is achievable with traditional independent learning algorithms.

— INSERT FIGURE 4 ABOUT HERE —

The comparative analysis of performance metrics reveals noteworthy patterns. When evaluating social welfare and fairness indicators between the standard and augmented algorithms, we observe that the augmented algorithm consistently outperforms the standard algorithm in terms of fairness. This improvement in fairness can be attributed to the algorithm's ability to alternate between the two pure strategy equilibria, thereby balancing the asymmetric payoff distribution inherent in each equilibrium. As expected, the social welfare remains unchanged across both implementations, since the sum of payoffs in either pure strategy Nash equilibrium is identical.

— INSERT FIGURE 5 ABOUT HERE —

Example 2. A 2×3 game with partial feedback and adaptive messages Consider the following game with two players and three actions each. The payoff structure is

	A_2	B_2	C_2
A_1	$0,0$	$5,1$	$1,5$
B_1	$1,5$	$0,0$	$5,1$
C_1	$5,1$	$1,5$	$0,0$

In this example we avoid the computation of a desirable correlated strategy a priori, and employ an adaptive message distribution approach rather than specifying a fixed probability distribution on messages. The adaptive message distribution allows the algorithm to discover efficient correlation patterns through the learning process itself. This approach is valuable in complex games where the optimal correlation structure is not immediately apparent from inspection of the payoff matrix. When analyzing the welfare metrics across a parameter grid, we observe that the average social welfare is improved, with the most substantial improvements occurring at intermediate iteration ranges. In parallel, the fairness metric shows a consistent and significant improvement across all parameter configurations of our grid.

— INSERT FIGURE 11 ABOUT HERE —

Example 3. A 3×2 game with full feedback Consider a game played by three players, where each player has two available actions. The payoff structure is characterized by the following payoff representation

A_3

	A_2	B_2
A_1	$0,0,0$	$0,0,1/2$
B_1	$10,10,0$	$0,0,0$

B_3

	A_2	B_2
A_1	$0,0,0$	$10,10,0$
B_1	$0,0,1/2$	$0,0,0$

We implement our algorithm with a message set $M = \{m_1, m_2\}$, where messages are public between players 1 and 2 but unobservable to player 3. Initially, we employ a stationary uniform probability distribution across messages. This information structure creates a perfect correlation that allows players 1 and 2 to learn their actions based on shared information that is unavailable to player 3. We evaluate the performance of this implementation across a grid of hyperparameters: we improve both social welfare and fairness compared to standard independent learning algorithms.

— INSERT FIGURE 12 ABOUT HERE —

We extend our investigation to examine the adaptive probability distribution over messages. In this variant, the distribution of messages evolves dynamically throughout the learning process in response to observed mixed strategies. The results from this implementation reveal an improvement in social welfare compared to both the baseline independent algorithm. However, fairness is not improved in the long run. This result reflects the specific objective function employed in our adaptive message distribution mechanism, which prioritizes social welfare maximization over fairness considerations.

— INSERT FIGURE 13 ABOUT HERE —

Example 4. A 3×3 game with partial feedback Consider the following game involving three players, each with three available actions. The strategic environment is defined by the following payoff structure

	A_2	B_2	C_2	
A_1	$0, 0, 10$	$0, 0, 0$	$-1, -1, 1$	
B_1	$5, 0, 0$	$0, 0, 0$	$-1, -1, 1$	A_3
C_1	$-1, -1, 1$	$-1, -1, 1$	$-1, -1, 1$	

	A_2	B_2	C_2	
A_1	$8, 8, 8$	$0, 0, 0$	$-1, -1, 1$	
B_1	$0, 0, 0$	$8, 8, 8$	$-1, -1, 1$	B_3
C_1	$-1, -1, 1$	$-1, -1, 1$	$-1, -1, 1$	

	A_2	B_2	C_2	
A_1	$0, 0, 0$	$0, 0, 0$	$-1, -1, 1$	
B_1	$0, 5, 0$	$0, 0, 10$	$-1, -1, 1$	C_3
C_1	$-1, -1, 1$	$-1, -1, 1$	$-1, -1, 1$	

We implement our algorithm with a message set $M = \{m_1, m_2\}$ and an initial uniform probability distribution. Messages are perfectly correlated and public between players 1 and 2, but unobservable to player 3. The results demonstrate mild improvements in both dimensions, specifically for some configurations of the hyperparameters, compared to standard independent learning algorithms.

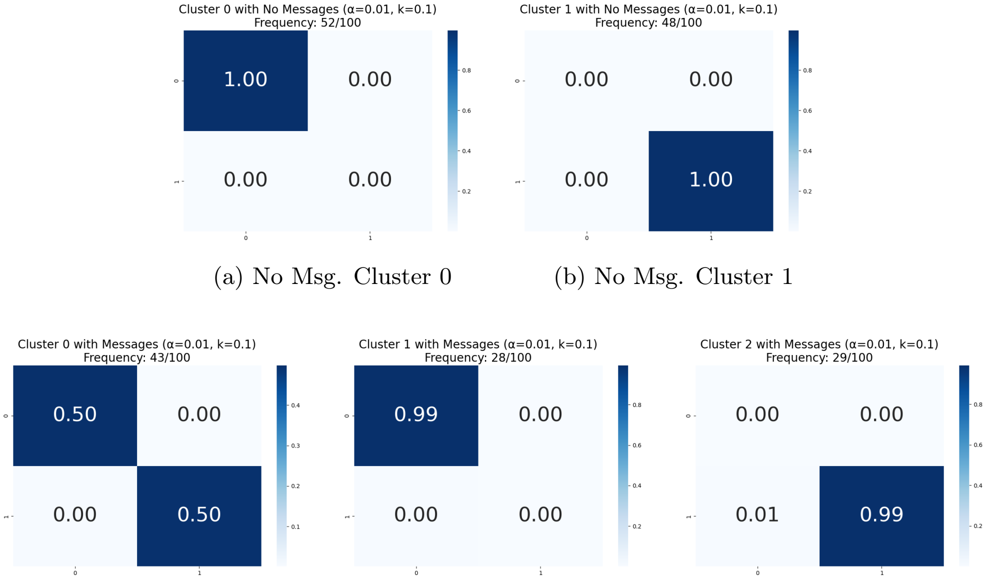
— INSERT FIGURE 15 ABOUT HERE —

We also investigate the impact of implementing an adaptive probability distribution over messages. The experimental results indicate that the adaptive approach does not improves social welfare compared to the stationary distribution case.

— INSERT FIGURE 17 ABOUT HERE —

9.2 Figures

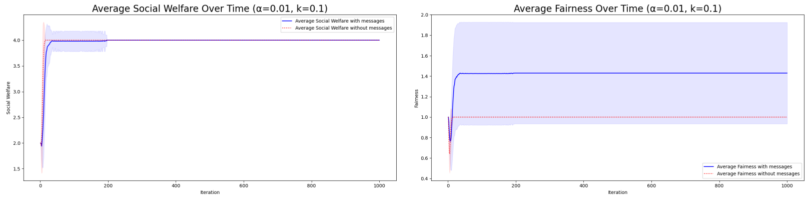
9.2.1 Coordination Game



(a) No Msg. Cluster 0 (b) No Msg. Cluster 1

(c) Msg. Cluster 0 (d) Msg. Cluster 1 (e) Msg. Cluster 2

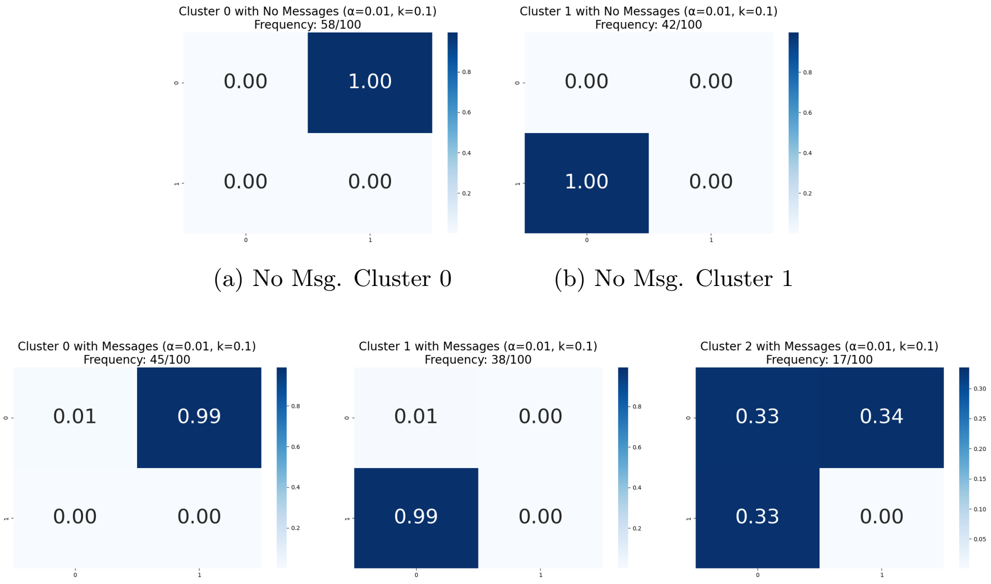
Figure 4: Comparison across scenarios: clusters of outcome frequency of play with no messages (a, b), and with messages (c, d, e)



(a) Social Welfare (b) Fairness

Figure 5: Comparison across scenarios: evolution of average social welfare and fairness with no messages (red) and with messages (blue).

9.2.2 Hawk-Dove Game



(a) No Msg. Cluster 0

(b) No Msg. Cluster 1

(c) Msg. Cluster 0

(d) Msg. Cluster 1

(e) Msg. Cluster 2

Figure 6: Comparison across scenarios: clusters of outcome frequency of play with no messages (a, b), and with messages (c, d, e)

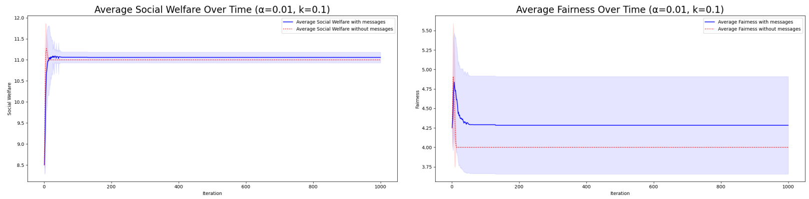


Figure 7: Comparison across scenarios: evolution of average social welfare and fairness with no messages (red) and with messages (blue).

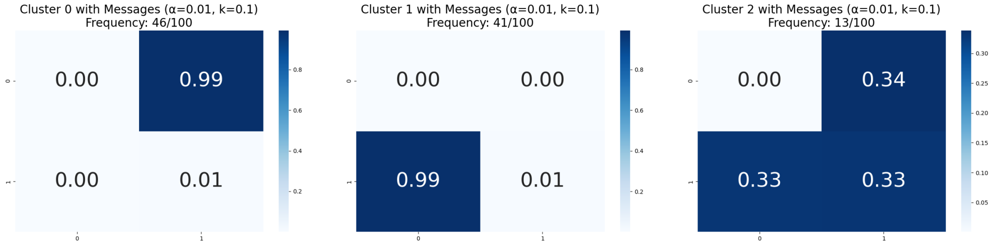


Figure 8: Clusters of outcome frequency of play with messages

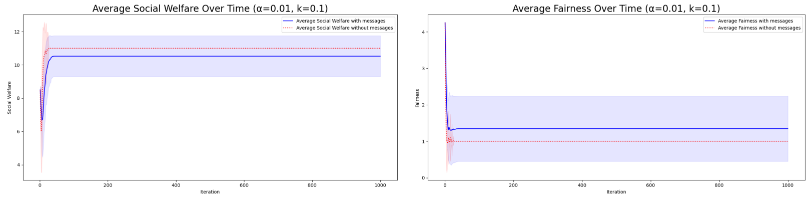


Figure 9: Comparison across scenarios: evolution of average social welfare and fairness with no messages (red) and with messages (blue).

9.2.3 2×3 game

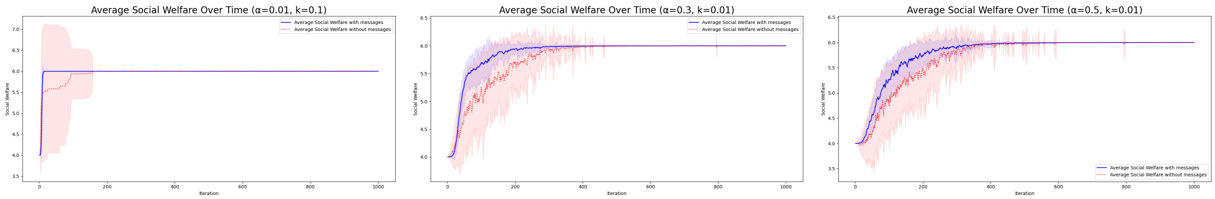


Figure 10: Comparison across scenarios: evolution of average social welfare with no messages (red) and with messages (blue). (adaptive message distribution)

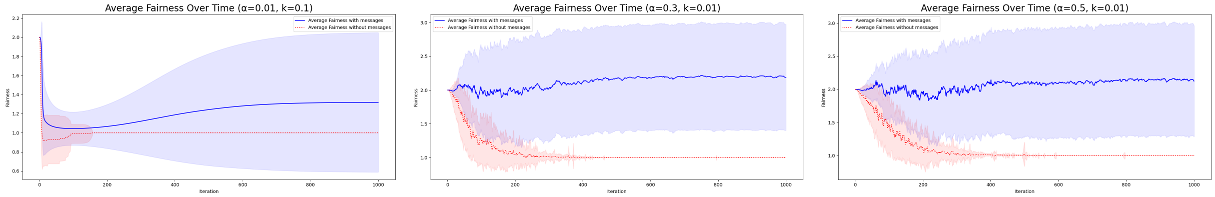


Figure 11: Comparison across scenarios: evolution of fairness with no messages (red) and with messages (blue). (adaptive message distribution)

9.2.4 3 × 2 game

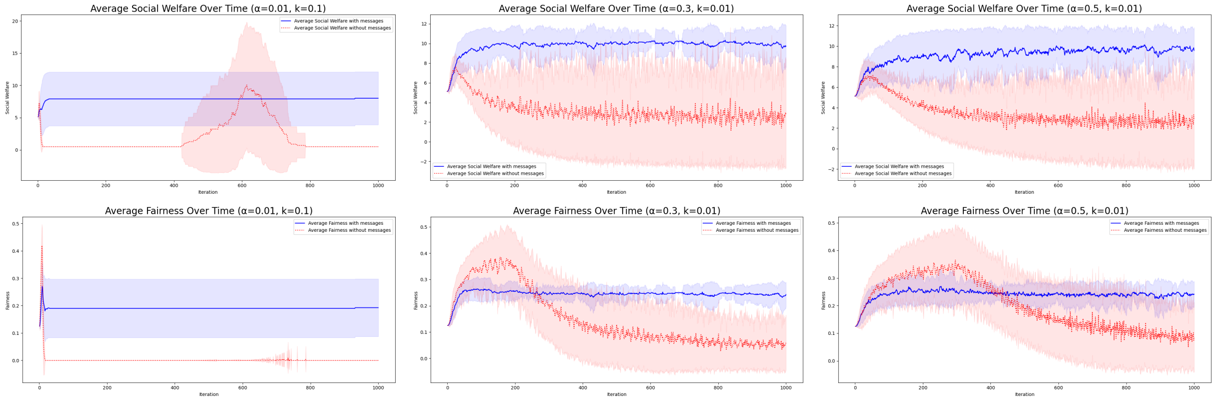


Figure 12: Comparison across scenarios: evolution of average social welfare and fairness with no messages (red) and with messages (blue). (stationary message distribution)

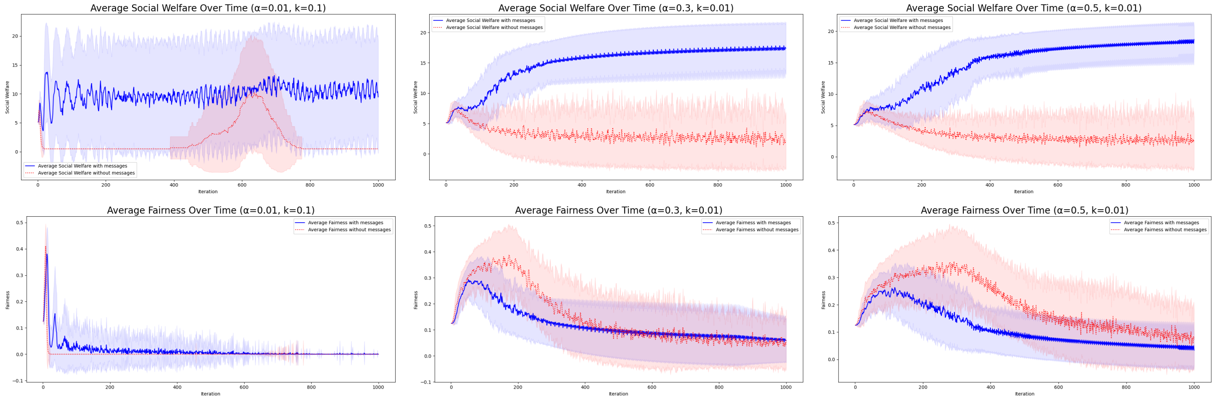


Figure 13: Comparison across scenarios: evolution of average social welfare and fairness with no messages (red) and with messages (blue). (adaptive message distribution)

9.2.5 3 × 3 game

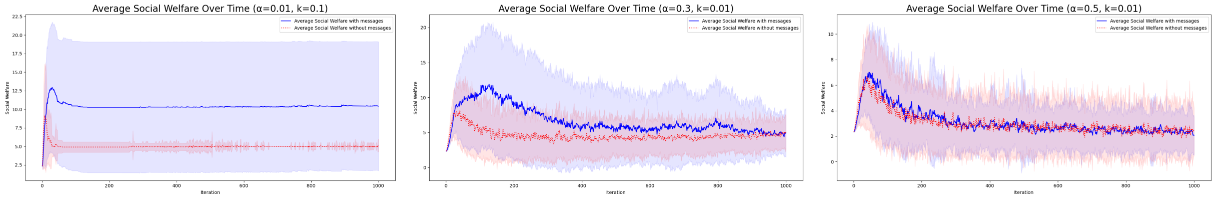


Figure 14: Comparison across scenarios: evolution of average social welfare with no messages (red) and with messages (blue). (stationary message distribution)

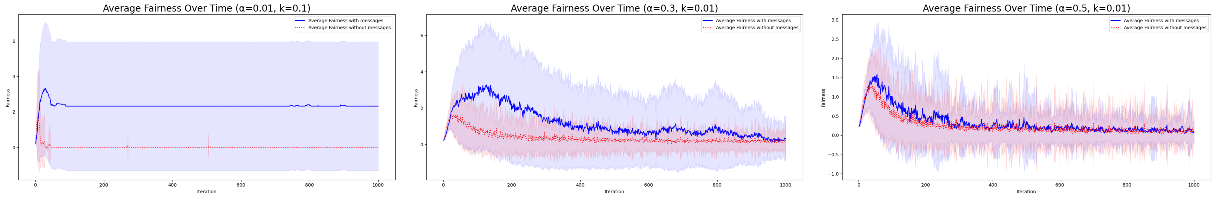


Figure 15: Comparison across scenarios: evolution of average fairness with no messages (red) and with messages (blue). (stationary message distribution)

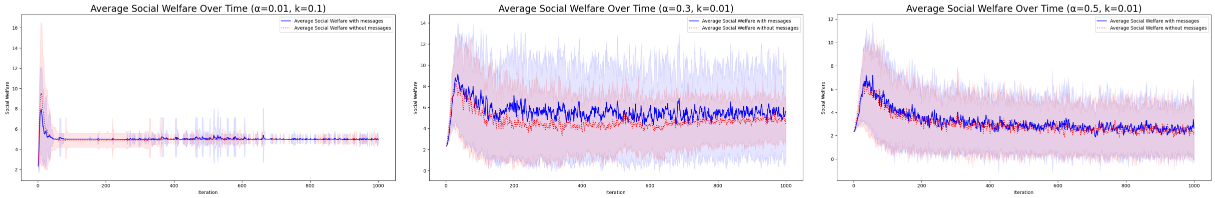


Figure 16: Comparison across scenarios: evolution of average social welfare with no messages (red) and with messages (blue). (adaptive message distribution)

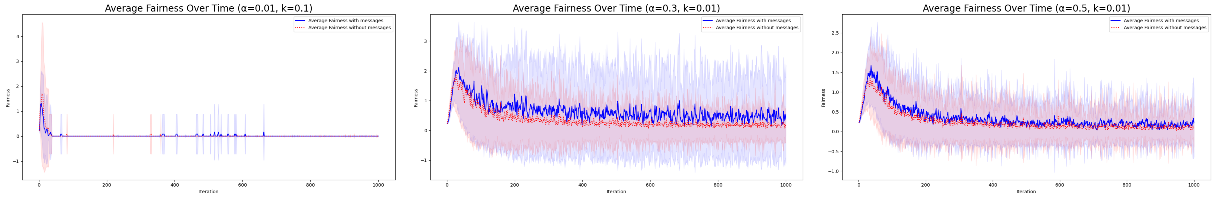


Figure 17: Comparison across scenarios: evolution of average fairness with no messages (red) and with messages (blue). (adaptive message distribution)