Boltzmann social learning with heterogeneous rationality

Sylvie Chaddad, Yezekael Hayel, Vineeth Varma, Nicolas Gast

Boltzmann social learning with heterogeneous rationality

Sylvie Chaddad[a,], Yezekael Hayel[a], Vineeth S. Varma[b], Nicolas Gast[c]

[a] *University of Avignon, LIA, Avignon, 84911, France*
[b] *University of Lorraine, CNRS, CRAN, Nancy, 54000, France*
[c] *University of Grenoble Alpes, INRIA, Grenoble, France*

Abstract

This paper analyzes a novel social learning model in which, at each discrete time step, agents with private preferences repeatedly select actions via a softmax (Boltzmann) rule, and update their preferences based on public observations of others' choices. This work addresses a critical gap by introducing *rational heterogeneity* through agent-specific rationality parameters γ_i. Unlike previous models, our approach accounts for the diverse ways individuals process social information by using a discrete-time deterministic mean-field approximation map. We establish fundamental equilibrium properties that were previously unexplored. In particular, we prove the existence of fixed points and show that, on complete graphs, every mean-field equilibrium is a consensus state, where all agents share identical preferences. We further derive sufficient conditions for the uniqueness of this equilibrium and its local asymptotic stability. Numerical simulations validate our theoretical findings and illustrate how rational heterogeneity and network structure interact to shape collective behavior in social learning systems.

Keywords: Multi-Agent Systems, Consensus equilibria, Stochastic Systems, Social Learning, Heterogeneous Rationality, Stability Analysis.

1. Introduction

Individuals often look to the experiences of others in their immediate networks (social and/or physical) when faced with behaviors that involve uncertain or unknown risks and outcomes. This process, known as *Social Learning (SL)*, can be formally described as a *collective behavior* phenomenon where agents form their opinions via the local exchange of information ([Cha04, JG10]), which typically occurs on social systems like online platforms through repeated observation and reaction to public actions such as up-votes, likes, subscriptions, or contributions ([TNB12]).

To capture these dynamics, researchers have employed frameworks such as interacting particle systems (IPS) and multi-agent systems (MAS). In probability theory, an IPS is a continuous-time Markov jump process defined on a configuration space and describes the collective behavior of stochastically interacting components ([Lig05]). Classical IPS examples include stochastic Ising models, voter models, contact processes, and linear systems, which have been used to model the spread of opinions within a population and to describe their dynamics in social systems ([SEM05]). MAS, on the other hand, is an algorithmic system composed of multiple interacting intelligent agents ([SLB09]), and are typically described by deterministic dynamics. Both frameworks consider large populations of units that interact with each other, and thus naturally model dynamics on social networks.

Current research integrates concepts from social networks and statistical physics within MAS frameworks. MAS models, techniques, and technologies have been used and have important potentialities for the study of social networks and the development of social network models. These approaches share structure and objectives, as they model

Email addresses: sylvie.chaddad@univ-avignon.fr (Sylvie Chaddad), yezekael.hayel@univ-avignon.fr (Yezekael Hayel), vineeth.satheeskumar-varma@univ-lorraine.fr (Vineeth S. Varma), nicolas.gast@inria.fr (Nicolas Gast)

individuals connected by relationships and are designed to achieve individual and collective goals ([FP11]). Particularly relevant to social learning are Boltzmann-type models that capture decision-making under uncertainty. For instance, [LT21] developed a Boltzmann-type kinetic model for interacting multi-agent systems where agents both interact and switch group labels via a state-dependent Markov jump process, yielding general kinetic equations for coupled interaction and label switch dynamics. While [GHB24a] analyzed independent Q learning with a Boltzmann exploration policy in MAS, and proposed a new discrete time approximation model that explicitly accounts for agents' update frequencies, whose dynamics fundamentally differ from earlier simplifications. Similarly, [PDPK15] introduced a Social Restricted Boltzmann Machine, an energy-based deep model, to predict human behavior in health social networks. This deep learning model effectively predicts activity levels by learning the complex interactions between these determinants in the network. Moreover, [SLA20] proposed a new multinomial logit learning model that integrates a neural network to learn a nonlinear utility representation within the standard logit choice framework.

These approaches align with social learning algorithms where information aggregation often follows Boltzmann-type distributions. For example, [BMS21] aggregates neighbors' information via log-linear models equivalent to softmax functions, leading to distinct behavioral clusters based on network structure and learning parameters. This social learning process leads to distinct clusters of behavior depending on memory, information transmission, learning about network outcomes, initial conditions, and long-range connections. [NMP17] further demonstrates how agents learn from the outcomes of their own and others' actions within their social networks.

The common practice of assuming *homogeneous* rationality in logit models of boundedly rational discrete choice can induce a *homogeneity bias* when the underlying population is in fact heterogeneous. In particular, Golman [Gol12] shows that forcing a single logit sensitivity (rationality) parameter to match the aggregate behavior of agents with heterogeneous sensitivities yields a *downward-biased* homogeneous parameter, so the population appears less rational than it truly is. Motivated by this observation, we propose a novel Boltzmann social-learning framework that explicitly incorporates *rational heterogeneity*, where each agent i has its own rationality parameter γ_i. Using a deterministic mean-field approximation (MFA) map, we analyze how the distribution of $\{\gamma_i\}$ shapes equilibrium existence, consensus structure, uniqueness, and local stability.

Table 1 compares our contribution with representative Boltzmann and social-learning works, and the last column highlights the novelty through keywords. We study a discrete-time social-learning mechanism in which agents repeatedly choose discrete actions under a logit rule while keeping *private* preference scores, and then update these scores using *public* observations of others' actions. Our main contribution is to develop a framework that explicitly captures rational heterogeneity: each agent has its own rationality parameter γ_i, capturing heterogeneous decision-making, and we analyze how this heterogeneity shapes equilibrium selection and stability.

We show that the dynamics remain in a compact invariant state space, establish the existence of a baseline equilibrium (the uniform fixed point), and show a key structural property on complete graphs: *every* mean-field equilibrium is necessarily a consensus state (all agents share the same preference vector). We then derive sufficient conditions for uniqueness and local asymptotic stability of this consensus equilibrium and validate the theory numerically, highlighting the interaction between heterogeneity and network structure. These results have not been previously established for heterogeneous Boltzmann-type social learning dynamics.

The remainder of this paper is structured as follows. Section 2 presents the stochastic model and its deterministic MFA. Section 3 establishes the existence and consensus property of fixed points, and derives conditions for stability, convergence, and uniqueness. Section 4 presents numerical simulations validating our theoretical results and exploring network effects. Finally, Section 5 concludes the paper and outlines future research directions.

Table 1: Comparison with representative Boltzmann and logit decision and social learning models.

Reference	Main contribution
Golman (2012) ([Gol12])	Showed that forcing a *homogeneous* logit and quantal response parameter on a population with heterogeneous rationality can lead to biased inference ("homogeneity bias").
Goll et al. (2024b) ([GHB24b])	Proposed a deterministic approximation in discrete time of incremental multi agent Boltzmann Q learning that explicitly accounts for agents' update frequencies.
Hussain and Belardinelli (2024) ([HB24])	Studied Q learning dynamics in competitive network games and characterize how competition, exploration rate, and network connectivity affect convergence and stability, highlighting the role of connectivity in delaying unstable behaviors.
Leonidov et al. (2025) ([LTV25])	Studied Boltzmann exploration in Q learning for the iterated Prisoner's Dilemma, emphasizing the balance between exploration (temperature) and strategic exploitation (discount factor) and the resulting learning regimes.
Wu et al. (2025) ([WDK+25])	Studied adaptive social learning and asocial learning in an immersive collective foraging task, behavioral and computational results show flexible mixing of local search and conditional social learning driven by individual success.
Taylor Davies et al. (2025) ([TDBL25])	Proposed a rational framework for selective social learning based on groups, where learners infer task relevant preferences of others and generalize using group signals, validated with behavioral experiments.
Novelty of "Boltzmann social learning with heterogeneous rationality" (Our work)	Social learning model with actions chosen using a Boltzmann rule. Using a mean-field approximation, we prove existence of fixed points which are *consensus* in states, and derive sufficient conditions for equilibrium uniqueness and local asymptotic stability.

2. Model

We consider a set of n agents indexed by $N = \{1, \dots, n\}$. At each discrete time t, agent i selects exactly one action from the discrete choices set $A = \{1, \dots, m\}$, denoted by $a_i(t) \in A$. Time is discretized, and agent i's decision at time $t \in \{1, \dots, T\}$ depends on an intrinsic preference vector $\theta_i(t) = (\theta_i^1(t), \dots, \theta_i^m(t))$ for all actions, where $\theta_i^k(t)$ represents the instantaneous score that agent i assigns to action k. Initial preferences are normalized such that $\theta_i^k(0) \in [0, 1]$ for all $i \in N$ and $k \in A$.

Given $\theta_i(t)$, agent i chooses action k randomly, with probability given by the softmax rule:

$$\mathbb{P}(a_i(t) = k) = \mathcal{F}_i^k(\theta_i(t)) = \frac{\exp(\gamma_i\, \theta_i^k(t))}{\sum_{l=1}^m \exp(\gamma_i\, \theta_i^l(t))}, \quad k \in A, \tag{1}$$

where $\gamma_i > 0$ is i's rationality parameter.

While actions are publicly observable in this complete graph setting, preferences remain private. Following discrete-time learning dynamics similar to [CMMS16] and [Mar07], preferences are updated via a discrete-time learning loop:

$$\theta_i^k(t+1) = \theta_i^k(t) + \varepsilon \sum_{j \in N} (\mathbb{1}_{\{a_j(t)=k\}} - \theta_i^k(t)), \tag{2}$$

where $0 < \varepsilon \leq \frac{1}{n}$ is the learning rate.

We analyze this stochastic process through its deterministic mean-field approximation, replacing random action selections with their expectations. This yields for all agents i and actions k, , the update map $G_i^k : \Omega \to \Omega$, where for any $\theta \in \Omega = [0, 1]^{n \times m}$:

$$G_i^k(\theta) \;=\; (1 - \varepsilon n)\, \theta_i^k \;+\; \varepsilon \sum_{j=1}^n \mathcal{F}_j^k(\theta_j), \tag{3}$$

with $0 < \varepsilon \leq 1/n$. The mean-field dynamics is generated by iterating $\theta(t+1) = G(\theta(t))$. The choice of the mean-field approximation over other stochastic modeling techniques, such as a Markov chain analysis, is motivated by analytical tractability for large populations. The MFA replaces $\mathbb{1}_{\{a_j(t)=k\}}$ in (2) by $\mathbb{E}[\mathbb{1}_{\{a_j(t)=k\}} \mid \theta(t)] = \mathcal{F}_j^k(\theta_j(t))$, yielding the deterministic map G. This enables stability analysis for large n and is consistent with successful mean field approaches developed for large-scale multi-agent learning dynamics ([HLL19, YLL+18]).

At equilibrium of the learning loop ($\theta_i^k(t+1) = \theta_i^k(t)$ for all i, k), the fixed-point condition becomes:

$$\sum_{j=1}^{n} \frac{\exp(\gamma_j \, \theta_j^k)}{\sum_{l=1}^{m} \exp(\gamma_j \, \theta_j^l)} = n \, \theta_i^k \quad \forall i \in N, \ \forall k \in A. \tag{4}$$

The map G is a continuous self-map of the nonempty, compact, convex set Ω. In the following section, we establish that G admits at least one fixed point and that any equilibrium must be a consensus state. Under specific conditions, this equilibrium is unique and locally asymptotically stable.

3. Analysis of Learning Dynamics

In this section, we show first preferences remain bounded, then establish fundamental properties of the system including the existence and consensus property of fixed points, derive conditions for uniqueness, stability, and global convergence, and analyze how these properties depend on agent rationality parameters.

3.1. Existence and Consensus Properties

We begin with the following lemma, which establishes a technical invariant property of the dynamics by proving that the learning dynamics preserve the boundedness of the agents' preference values, ensuring that the state remains within the normalized interval $[0, 1]$.

Lemma 1. *The learning dynamics* (2) *preserve the boundedness of preferences, i.e.* $\theta_i^k(t) \in [0, 1]$ *for all* i, k *and* $t \geq 0$.

Proof. By induction on t. The base case $t = 0$ holds by assumption. For the inductive step, fix i, k and let $x = \theta_i^k(t) \in [0, 1]$, $y = \sum_{j=1}^{n} \mathbb{1}_{\{a_j(t)=k\}} \in \{0, \ldots, n\}$. The update rule gives:

$$\theta_i^k(t+1) = (1 - \varepsilon n)x + \varepsilon y.$$

Since $\varepsilon \leq 1/n$ ensures $1 - \varepsilon n \geq 0$, and $0 \leq x \leq 1$, $0 \leq y \leq n$, we have:

$$0 = (1 - \varepsilon n) \cdot 0 + \varepsilon \cdot 0 \leq \theta_i^k(t+1) \leq (1 - \varepsilon n) \cdot 1 + \varepsilon \cdot n = 1.$$

Thus $\theta_i^k(t+1) \in [0, 1]$, completing the induction. □

This boundedness ensures the mean-field map G is a continuous self-map on the compact convex set $\Omega = [0, 1]^{n \times m}$. In other words, it ensures the stability of the learning process and guarantees numerical robustness.

We now state a first equilibrium existence result. The following proposition proves that $G : \Omega \to \Omega$ has at least one fixed point, the uniform point.

Proposition 3.1. *The uniform point* $c_{unif} = (1/m, \ldots, 1/m)$ *is a fixed point of* G.

Proof. At the uniform point, all actions have equal preference $1/m$. For any agent j, the softmax probability becomes:

$$\mathcal{F}_j^k(c_{\text{unif}}) = \frac{e^{\gamma_j \cdot (1/m)}}{\sum_{\ell=1}^{m} e^{\gamma_j \cdot (1/m)}} = \frac{e^{\gamma_j/m}}{m \cdot e^{\gamma_j/m}} = \frac{1}{m}.$$

Applying the mean-field map, for any i, k,

$$G_i^k(c_{\text{unif}}) = (1 - \varepsilon n)\frac{1}{m} + \varepsilon \sum_{j=1}^{n} \mathcal{F}_j^k(c_{\text{unif}}) = \frac{1}{m} = c_{\text{unif}}^k.$$

Therefore, $(c_{\text{unif}}, \ldots, c_{\text{unif}})$ is a fixed point of G. □

The existence of this fixed point ensures that the system always has at least one equilibrium: a consensus state in which all actions are equally preferred. The following results establish a fundamental property of the system at equilibrium: all agents must share an identical preference vector (i.e., any fixed point of G must be a consensus point).

Proposition 3.2 (Consensus at fixed points)**.** *Let $\theta^* \in \Omega$ be a fixed point of the map G, i.e., $G(\theta) = \theta$. Then, all agents share identical preference vectors. Specifically, there exists $c = (c_1, \ldots, c_m)$ in the probability simplex $\Delta^m = \{x \in [0,1]^m / \sum_{k=1}^m x^k = 1\}$ such that:*

$$\theta_i^* = c \quad \forall i \in N$$

Furthermore, this consensus vector satisfies:

$$c^k = \theta_i^{*k} = \frac{1}{n} \sum_{j=1}^{n} \frac{\exp(\gamma_j c^k)}{\sum_{\ell=1}^{m} \exp(\gamma_j c^\ell)} \quad \text{for all } k \in A. \tag{5}$$

With $\sum_{k=1}^m c^k = 1$. Hence, any fixed point of G is a consensus equilibrium.

Proof. From the fixed point condition $G(\theta^*) = \theta^*$, for any $i \in N$ and $k \in A$:

$$\theta_i^{*k} = (1 - \varepsilon n)\theta_i^{*k} + \varepsilon \sum_{j=1}^{n} \mathcal{F}_j^k(\theta_j^*)$$

Solving,

$$\theta_i^{*k} - (1 - \varepsilon n)\theta_i^{*k} = \varepsilon \sum_{j=1}^{n} \mathcal{F}_j^k(\theta_j^*)$$

$$\varepsilon n\, \theta_i^{*k} = \varepsilon \sum_{j=1}^{n} \mathcal{F}_j^k(\theta_j^*)$$

$$\theta_i^{*k} = \frac{1}{n} \sum_{j=1}^{n} \mathcal{F}_j^k(\theta_j^*), \tag{6}$$

The right-hand side of (6) is independent of the agent index i. Therefore, for each action k, θ_i^{*k} are identical across all agents i. Denote this common value by c^k. Therefore $\theta_i^* = c$ for all $i \in N$.

Summing over $k \in A$:

$$\sum_{k=1}^{m} c^k = \frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{m} \mathcal{F}_j^k(c) = \frac{1}{n} \sum_{j=1}^{n} 1 = 1,$$

so $c \in \Delta^m$. Substituting $\theta_j^* = c$ into (6) proves the equilibrium condition. □

In practical terms, this proves that social learning leads to agreement. In real-world terms, it shows how heterogeneous individuals reach a common decision even when members start with different personal preferences.

An immediate consequence of Proposition 3.2 is the following corollary, which explicitly establishes that non-consensus equilibria cannot exist.

Corollary 1. *Every fixed point of G is in consensus of preference across agents (i.e., there are no non-consensus fixed points).*

Proof. Assume that there exists a fixed point $\tilde{\theta} \in \Omega$ that is not a consensus. Then there exist $p \neq q$ and $k \in A$ with $\tilde{\theta}_p^k \neq \tilde{\theta}_q^k$. However, the fixed-point implies $\forall\, i$ and $k \in A$:

$$\tilde{\theta}_i^k = \frac{1}{n} \sum_{j=1}^{n} \mathcal{F}_j^k(\tilde{\theta}_j),$$

where the right-hand side is independent of i. Hence $\tilde{\theta}_p^k = \tilde{\theta}_q^k$ for all k, a contradiction. □

The results of Section 3.1 ensure that preferences remain bounded and that any equilibrium of the mean-field dynamics is a consensus. We now turn to the questions of uniqueness and stability of this consensus equilibrium in the presence of heterogeneous rationality, which is analyzed in the following Section 3.2.

3.2. Uniqueness and stability

Given that agents may have different levels of rationality (different γ_j), we study how this heterogeneity in agent rationality affects the uniqueness, stability, and convergence properties of fixed points. This reflects real-world scenarios where agents may have different levels of decision-making.

The following theorem bounds the Jacobian's spectral radius, yielding local stability for the consensus fixed point.

Theorem 1. *Let $\theta^* = (c, \dots, c)$ be a consensus fixed point. If $\sum_{j=1}^{n} \gamma_j < 2n$, then all eigenvalues of $DG(\theta^*)$ are strictly inside the unit disk, hence θ^* is locally asymptotically stable.*

Proof. We analyze the Jacobian of the mean-field map G. Since $\mathcal{F}_j$ depends only on θ_j,

$$\frac{\partial \mathcal{F}_j^k}{\partial \theta_r^s} = \begin{cases} \gamma_j\big(\delta_{ks}\,\mathcal{F}_j^k - \mathcal{F}_j^k \mathcal{F}_j^s\big), & r = j, \\ 0, & r \neq j. \end{cases}$$

At the consensus point θ^*, $p^{(j)} = \mathcal{F}_j(\theta_j^*) = \mathcal{F}_j(c) \in \Delta^m$ and the blocks of the Jacobian is are:

$$H^{(j)} \;=\; p_k^{(j)}\,\delta_{ks} \;-\; p_k^{(j)}\,p_s^{(j)}, = \;\mathrm{Diag}(p^{(j)}) - p^{(j)}(p^{(j)})^{\top}$$

Hence the Jacobian $J \in \mathbb{R}^{nm\times nm}$ is:

$$J_{(i,k),(r,s)} = (1 - \varepsilon n)\,\delta_{ir}\delta_{ks} + \varepsilon\,\gamma_r\,\mathcal{F}_r^k(\theta_r)\,(\delta_{ks} - \mathcal{F}_r^s(\theta_r)).$$

Then the Jacobian at θ^* has the form

$$J \;=\; (1 - \varepsilon n)\,I_{nm} \;+\; \varepsilon\,K$$

with

$$K \;=\; \begin{pmatrix} \gamma_1 H^{(1)} & \gamma_2 H^{(2)} & \cdots & \gamma_n H^{(n)} \\ \gamma_1 H^{(1)} & \gamma_2 H^{(2)} & \cdots & \gamma_n H^{(n)} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_1 H^{(1)} & \gamma_2 H^{(2)} & \cdots & \gamma_n H^{(n)} \end{pmatrix}.$$

$$I_{nm} = I_n \otimes I_m, \qquad (I_{nm})_{(i,k),(r,s)} = \delta_{ir}\,\delta_{ks},$$

All block rows are identical, so $\mathrm{rank}(K) \le m$.
Let the m-dimensional subspace of block-constant vectors, $\mathcal{S} = \{v \in \mathbb{R}^{nm} \,:\, v_1 = \cdots = v_n\}$, where $v_i \in \mathbb{R}^m$ is the i-th agent block. For any $v = (v_1, \dots, v_n)$,

$$Kv \;=\; \begin{pmatrix} \sum_{r=1}^{n} \gamma_r H^{(r)} v_r \\ \vdots \\ \sum_{r=1}^{n} \gamma_r H^{(r)} v_r \end{pmatrix} \in \mathcal{S}.$$

Thus $\mathrm{Im}(K) \subseteq \mathcal{S}$. In particular, for $v = (w, \dots, w) \in \mathcal{S}$, $Kv = (T(w), \dots, T(w)),$, where $T(w) = \sum_{r=1}^{n} \gamma_r H^{(r)} w$. Therefore, the nonzero eigenvalues of K are exactly the eigenvalues of the $m \times m$ map T. Each $H^{(j)}$ is symmetric positive semidefinite with $H^{(j)}\mathbf{1}_m = 0$ (rows sum to zero), so it has at least one zero eigenvalue and $\mathrm{rank}(H^{(j)}) \le m - 1$. Let $H = \mathrm{Diag}(p) - pp^{\top}$ with $p \in \Delta_m$. Its entries are:

$$H_{kk} = p_k(1 - p_k), \qquad H_{ks} = -p_k p_s \quad (s \neq k).$$

For each row k, the Gershgorin center and radius are

$$c_k = H_{kk} = p_k(1 - p_k),$$

$$R_k = \sum_{s \neq k} |H_{ks}| = \sum_{s \neq k} p_k p_s = p_k(1 - p_k).$$

Hence every eigenvalue λ of H lies in:

$$\bigcup_k \{ z \in \mathbb{C} : |z - c_k| \leq R_k \} \implies \lambda_{\max}(H) \leq \max_k (c_k + R_k) = \max_k (2\, p_k(1 - p_k)) \leq \tfrac{1}{2},$$

Apply to each block, for $H^{(j)} = \mathrm{Diag}(p^{(j)}) - p^{(j)}(p^{(j)})^\top$ with $p^{(j)} = \mathcal{F}_j(c) \in \Delta_m$:

$$\lambda_{\max}(H^{(j)}) \leq \tfrac{1}{2} \quad \implies \quad \|H^{(j)}\|_2 \leq \tfrac{1}{2}.$$

For symmetric matrices, the spectral norm $\|H^{(j)}\|_2$ equals the largest absolute eigenvalue:

$$\|H^{(j)}\|_2 = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } H^{(j)}\}.$$

Since all eigenvalues are nonnegative, this simplifies to $\|H^{(j)}\|_2 = \lambda_{\max}(H^{(j)})$.
Let $T = \sum_{r=1}^n \gamma_r H^{(r)}$ with $\gamma_r \geq 0$. For any unit vector u,

$$u^\top T u = \sum_{r=1}^n \gamma_r\, u^\top H^{(r)} u \leq \sum_{r=1}^n \gamma_r\, \lambda_{\max}(H^{(r)}) \leq \tfrac{1}{2} \sum_{r=1}^n \gamma_r.$$

Because each $H^{(r)}$ is PSD and $\|H^{(r)}\|_2 \leq \frac{1}{2}$.

Taking the supremum over $\|u\|_2 = 1$ yields:

$$\lambda_{\max}(T) = \|T\|_2 \leq \tfrac{1}{2} \sum_{r=1}^n \gamma_r,$$

i.e. $\mu_{\max}(T) \leq \frac{1}{2} \sum_{r=1}^n \gamma_r$. Since $J = (1 - \varepsilon n) I_{nm} + \varepsilon K$, each eigenvalue of J has the form $\lambda = 1 - \varepsilon n + \varepsilon\mu$, with μ an eigenvalue of K.

On $\mathcal{S}^\perp$, $\mu = 0$ so $\lambda = 1 - \varepsilon n \in [0, 1)$ for $0 < \varepsilon \leq 1/n$.

On $\mathcal{S}$, $0 \leq \mu \leq \frac{1}{2} \sum_{r=1}^n \gamma_r$, hence $\lambda_{\max} \leq 1 - \varepsilon n + \varepsilon \frac{1}{2} \sum_{r=1}^n \gamma_r < 1$ whenever $\sum_{r=1}^n \gamma_r < 2n$.

Moreover, because K is positive semi-definite, all $\mu \geq 0$ and hence all $\lambda \geq 1 - \varepsilon n \geq 0$, so $|\lambda| < 1$ holds for every eigenvalue. Therefore, the spectral radius $\rho(J) < 1$, and the consensus fixed point θ^* is locally asymptotically stable by the linearized stability principle for discrete time maps ([AG14, Ela05]).

□

The following theorem provides that under the same condition, the mean-field map becomes a contraction, implying uniqueness and global convergence to equilibrium.

Definition 1. *A fixed point θ^* of the system $\theta(t+1) = G(\theta(t))$ is globally exponentially attractive if there exist constants $C \geq 1$ and $\rho \in (0, 1)$ such that for all $\theta(0) \in \Omega$ and all $t \geq 0$:*

$$\|\theta(t) - \theta^*\|_\infty \leq C\rho^t \|\theta(0) - \theta^*\|_\infty.$$

Theorem 2. *If $\sum_{j=1}^n \gamma_j < 2n$, then:*

1. *G is a contraction on $(\Omega, \|\cdot\|_\infty) \implies$ G admits a unique fixed point, $\theta^* \in \Omega$,*
2. *θ^* is globally exponentially attractive.*

Proof. Consider the complete metric space $(\Omega, \|\cdot\|_\infty)$, where $\Omega = [0,1]^{n\times m}$ is closed and $G(\Omega) \subseteq \Omega$.

For $\theta, \theta' \in \Omega$ and $i \in N$, $k \in A$, we have:

$$\left|G_i^k(\theta) - G_i^k(\theta')\right| = \left|(1-\varepsilon n)(\theta_i^k - \theta_i'^k) + \varepsilon \sum_{j=1}^{n} (\mathcal{F}_j^k(\theta_j) - \mathcal{F}_j^k(\theta_j'))\right| \leq (1-\varepsilon n)\left|\theta_i^k - \theta_i'^k\right| + \varepsilon \sum_{j=1}^{n} \left|\mathcal{F}_j^k(\theta_j) - \mathcal{F}_j^k(\theta_j')\right|.$$

where $F^k(\theta_j)$ is the homogeneous softmax.

Being Lipschitz continuous under $\|\cdot\|_\infty$, we compute the gradient to bound the difference via the mean value theorem. The gradient of F^k is:

$$\nabla_{\theta_j} \mathcal{F}^k = \gamma_j \mathcal{F}^k(\theta_j)\,(e_k - \mathcal{F}(\theta_j))$$

which is the k-th row of the Jacobian matrix of the softmax function ([GP17]). Here, $\mathbf{e}_k$ is the standard basis vector. Its L1-norm is:

$$\|\nabla_{\theta_j} \mathcal{F}^k\|_1 = \gamma\, \mathcal{F}^k(\theta_j)\, \|e_k - \mathcal{F}(\theta_j)\|_1.$$

Compute:

$$\left\|\mathbf{e}_k - \mathcal{F}(\theta_j)\right\|_1 = \left|1 - \mathcal{F}^k(\theta_j)\right| + \sum_{l\neq k} \left|0 - \mathcal{F}^l(\theta_j)\right| = (1 - \mathcal{F}^k(\theta_j)) + (1 - \mathcal{F}^k(\theta_j)) = 2\,(1 - \mathcal{F}^k(\theta_j)).$$

Thus:

$$\|\nabla_{\theta_j} \mathcal{F}^k\|_1 = 2\gamma_j \mathcal{F}^k(\theta_j)(1 - \mathcal{F}^k(\theta_j)).$$

The maximum of $\mathcal{F}^k(1-\mathcal{F}^k)$ is $1/4$, so:

$$\|\nabla_{\theta_j} \mathcal{F}^k\|_1 \leq 2\gamma_j \cdot \tfrac{1}{4} = \gamma_j/2.$$

By the mean value theorem:
$\left|\mathcal{F}^k(\theta_j) - \mathcal{F}^k(\theta_j')\right| \leq (\gamma_j/2)\,\|\theta_j - \theta_j'\|_\infty \leq (\gamma_j/2)\,\|\theta - \theta'\|_\infty$. Therefore:

$$\left|G_i^k(\theta) - G_i^k(\theta')\right| \leq (1-\varepsilon n)\|\theta - \theta'\|_\infty + \varepsilon \sum_{j=1}^{n} \frac{\gamma_j}{2}\|\theta - \theta'\|_\infty \leq \left(1 - \varepsilon n + \varepsilon \frac{1}{2}\sum_{j=1}^{n} \gamma_j\right)\|\theta - \theta'\|_\infty.$$

This holds for all i and k, so:

$$\|G(\theta) - G(\theta')\|_\infty \leq \left(1 - \varepsilon n + \varepsilon \frac{1}{2}\sum_{j=1}^{n} \gamma_j\right)\|\theta - \theta'\|_\infty.$$

Let $c = 1 - \varepsilon n + \varepsilon \frac{1}{2}\sum_{j=1}^{n} \gamma_j$. If $\sum_{j=1}^{n} \gamma_j < 2n$, then $c < 1$. Thus, G is a contraction on Ω with contraction constant c. By the Banach fixed-point theorem, G admits a unique fixed point $\theta^* \in \Omega$ and the contraction ensures the global convergence to the unique fixed point θ^* from any initial condition $\theta(0) \in \Omega$.

Moreover, since G is c–Lipschitz and $G(\theta^*) = \theta^*$,

$$\|\theta(t+1) - \theta^*\|_\infty = \|G(\theta(t)) - G(\theta^*)\|_\infty \leq c\,\|\theta(t) - \theta^*\|_\infty.$$

By induction,

$$\|\theta(t) - \theta^*\|_\infty \leq c^t\,\|\theta(0) - \theta^*\|_\infty \quad \forall t \geq 0,$$

which establishes global exponential attractivity with $c \in (0,1)$. □

Remark 1. *Lemma 3.1 established that the uniform point $\mathbf{c}_{unif} = (1/m, \ldots, 1/m)$ is a fixed point of G. Under the condition $\sum_{j=1}^{n} \gamma_j < 2n$, the contraction mapping guarantees that G admits exactly one fixed point. Therefore, the unique fixed point θ^* is the uniform point $\mathbf{c}_{unif}$.*

The following Theorem shows that when all agents prefer the same action, the fixed points become locally exponentially stable as all rationality parameters tend to infinity.

Theorem 3. *Let $k \in \{1, \dots, m\}$ and let θ^* be a fixed point of G, where all agents concentrate on the same action k, i.e., $\theta_i^* = e_k$ for all $i \in N$, where e_k is the k-th standard basis vector in $\mathbb{R}^m$. Then, for sufficiently high rationality $\gamma_j \to \infty$ for all $j \in N$, all fixed points θ^* of this form are locally asymptotically stable.*

Proof. Consider the fixed point $\theta^* = (e_k, \dots, e_k)$ where all agents prefers action k. For each agent j (i.e., $\theta_j^{*k} = 1$ and $\theta_j^{*\ell} = 0$ for $\ell \neq k$), we compute the softmax probabilities at θ_j^*:

$$\mathcal{F}_j^k(\theta_j^*) = \frac{e^{\gamma_j \cdot 1}}{e^{\gamma_j \cdot 1} + \sum_{\ell \neq k} e^{\gamma_j \cdot 0}} = \frac{e^{\gamma_j}}{e^{\gamma_j} + (m-1)},$$

$$\mathcal{F}_j^\ell(\theta_j^*) = \frac{e^{\gamma_j \cdot 0}}{e^{\gamma_j} + (m-1)} = \frac{1}{e^{\gamma_j} + (m-1)} \quad (\ell \neq k).$$

Define the normalization constant $Z_j = e^{\gamma_j} + (m-1)$ and let:

$$a_j = \frac{1}{Z_j}, \quad b_j = \frac{e^{\gamma_j}}{Z_j}.$$

The probability vector becomes:

$$p^{(j)} = \mathcal{F}_j(e_k) = (a_j, \dots, a_j, \underbrace{b_j}_{k\text{-th}}, a_j, \dots, a_j),$$

As $\gamma_j \to \infty$, we have:

$$b_j = \frac{1}{1 + (m-1)e^{-\gamma_j}} \to 1, \quad a_j = \frac{e^{-\gamma_j}}{1 + (m-1)e^{-\gamma_j}} \to 0,$$

so $p^{(j)} \to e_k$.

The Jacobian of G at θ^* has the structure:

$$J \;=\; (1 - \varepsilon n)\, I_{nm} \;+\; \varepsilon\, K,$$

where I_{nm} is the $nm \times nm$ identity and K is the block matrix with blocks $\gamma_j H^{(j)}$ and $H^{(j)} = \mathrm{Diag}(p^{(j)}) - p^{(j)}(p^{(j)})^\top$.

Using Taylor expansion $(1+x)^{-1} = 1 - x + O(x^2)$ with $x = (m-1)e^{-\gamma_j}$:

$$b_j = 1 - (m-1)e^{-\gamma_j} + O(e^{-2\gamma_j}),$$
$$a_j = e^{-\gamma_j} + O(e^{-2\gamma_j}).$$

Since

$$H_{rs}^{(j)} = p_r^{(j)} \delta_{rs} - p_r^{(j)} p_s^{(j)},$$

The entries of $H^{(j)}$ satisfy:

- **Diagonal entries:**

$$H_{kk}^{(j)} = b_j(1 - b_j) = (m-1)e^{-\gamma_j} + O(e^{-2\gamma_j}),$$
$$H_{\ell\ell}^{(j)} = a_j(1 - a_j) = e^{-\gamma_j} + O(e^{-2\gamma_j}), \quad \ell \neq k.$$

- **Off-diagonal entries:**

$$H_{k\ell}^{(j)} = -b_j a_j = -e^{-\gamma_j} + O(e^{-2\gamma_j}), \quad \ell \neq k,$$
$$H_{\ell k}^{(j)} = -a_j b_j = -e^{-\gamma_j} + O(e^{-2\gamma_j}), \quad \ell \neq k,$$
$$H_{\ell r}^{(j)} = -a_j^2 = O(e^{-2\gamma_j}), \quad \ell \neq r, \ell, r \neq k.$$

All entries are of order $O(e^{-\gamma_j})$, hence $\|H^{(j)}\| = O(e^{-\gamma_j})$

Consequently:

$$\Rightarrow \quad \gamma_j \|H^{(j)}\| = O(\gamma_j e^{-\gamma_j}) \to 0 \quad \text{as} \quad \gamma_j \to \infty.$$

Since each block $\gamma_j H^{(j)} \to 0$, we have $K \to 0$, and:

$$J \to (1 - \varepsilon n) I_{nm}.$$

Given that $0 < \varepsilon \leq \frac{1}{n}$, we have $0 \leq 1 - \varepsilon n < 1$, so the spectral radius $\rho(J) < 1$. Therefore, θ^* is locally asymptotically stable for sufficiently large γ_j. □

3.3. Convergence result

In this section, we show that the mean field model is a good approximation of the stochastic behavior for large n. Our analysis is done in the case where the mean field model G is a contraction. Also, to simplify the analysis, we consider a learning rate $\varepsilon = 1/n$. With this parameters, our result shows that $\mathbb{E}\left[\|\theta(t) - \theta^*\|_\infty\right]$ is of order $O(1/\sqrt{n})$ when n is large. As indicated in the proof, using a smaller learning rate would lead to a more accurate error rate of $\varepsilon\sqrt{n}$.

Theorem 4. *Assume that $\sum_{j=1}^n \gamma_j < 2n$. Then, for any $\varepsilon \in (0, 1/n)$:*

$$\lim_{t\to\infty} \mathbb{E}\left[\|\theta(t) - \theta^*\|_\infty\right] \leq \frac{\sqrt{m/n}}{1 - \sum_{j=1}^n \gamma_j / 2n}. \tag{7}$$

Proof. Let $M^k(t)$ be a random variable defined as for any action $k \in A$:

$$M^k(t) := \varepsilon \sum_{j=1}^n \left(\mathbb{1}_{\{a_j(t)=k\}} - \mathcal{F}_j^k(\theta_j(t)) \right).$$

Conditioned on $\theta(t)$, the term $\sum_{j=1}^n \mathbb{1}_{\{a_j(t)=k\}}$ is a sum of n independent Bernoulli random variables. This shows that:

$$\mathrm{Var}\left[M^k(t) \mid \theta(t)\right] = \sum_{j=1}^n \mathrm{Var}\left[\mathbb{1}_{\{a_j(t)=k\}} \mid \theta(t))\right] = \sum_{j=1}^n \mathcal{F}_j^k(\theta_j(t))(1 - \mathcal{F}_j^k(\theta_j(t))) \leq n.$$

As $M^k(t)$ is zero-mean, this implies that:

$$\mathbb{E}\left[|M^k(t)|^2\right] = \mathbb{E}\left[\mathrm{Var}\left[M^k(t) \mid \theta(t)\right]\right] \leq n.$$

By using Jensen's inequality and the fact that $\|M(t)\|_\infty^2 \leq \sum_{k=1}^m |M^k(t)|^2$, this implies that:

$$\mathbb{E}\left[\|M(t)\|_\infty\right] \leq \sqrt{\mathbb{E}\left[\|M(t)\|_\infty^2\right]} \leq \sqrt{\sum_{k=1}^m \mathbb{E}\left[|M^k(t)|^2\right]} \leq \sqrt{mn}.$$

Equation (2) can be rewritten as:

$$\theta(t+1) = G(\theta(t)) + \varepsilon M(t).$$

Let $\nu(t+1)$ be the deterministic trajectory, defined as:

$$\nu(t+1) = G(\nu(t)).$$

By using that G is a c-contraction with $c = 1 - \varepsilon n + \frac{\varepsilon}{2} \sum_{j=1}^n \gamma_j$, we have:

$$\|\theta(t+1) - \nu(t+1)\|_\infty \leq \|G(\theta(t)) - G(\nu(t)\|_\infty + \varepsilon \|M(t)\|_\infty$$

Taking expectations and using the bound on $\mathbb{E}[||M(t)||]$ yields

$$\mathbb{E}[||\theta(t+1) - \nu(t+1)||] \leq c\,\mathbb{E}[||\theta(t) - \nu(t)||] + \varepsilon\sqrt{mn}.$$

Iterating this inequality and using $\sum_{\ell=0}^{t} c^\ell \leq \frac{1}{1-c}$ (since $c < 1$) gives

$$\sup_{t\geq 0} \mathbb{E}[||\theta(t) - \nu(t)||] \leq \frac{\varepsilon\sqrt{mn}}{1-c}.$$

Moreover, since G is a contraction, $\nu(t) \to \theta^*$ as $t \to \infty$, hence

$$\limsup_{t\to\infty} \mathbb{E}[||\theta(t) - \theta^*||] \leq \frac{\varepsilon\sqrt{mn}}{1-c}.$$

Replacing c by its value $1 - \varepsilon n + \frac{\varepsilon}{2}\sum_{j=1}^{n}\gamma_j$ implies that:

$$\frac{\varepsilon\sqrt{mn}}{1-c} = \frac{\varepsilon\sqrt{mn}}{n\varepsilon - \frac{\varepsilon}{2}\sum_{j=1}^{n}\gamma_j} = \frac{\sqrt{m/n}}{1 - \frac{1}{2n}\sum_{j=1}^{n}\gamma_j},$$

where on the second line we divided the numerator and denominator by $n\varepsilon$. □

The previous result shows that $\mathbb{E}[||\theta(t) - \theta^*||_\infty]$ is of order $O(1/\sqrt{n})$. The next proposition shows that $||\mathbb{E}[\theta(t)] - \theta^*||_\infty = O(1/n)$. In other words, if $\theta(t)$ is at a distance $O(1/\sqrt{n})$ from θ^*, its expectation $\mathbb{E}[\theta(t)]$ is at a distance $O(1/n)$ from θ^*. This is much smaller, which justifies the mean-field approximation. To keep the proof simple, we only prove the result for $\varepsilon = 1/n$. This allows us to apply the results of [GLM18] directly. We believe that the approach of [GLM18] could be adapted to the case $\varepsilon < 1/n$, but this would likely require a more complex approach to deal with the heterogeneity of the values θ_i for $i \neq j$ (for instance, by using the framework of [AG22]). As the proof of such a result is quite technical, we left a detailed proof of this result for future work.

Proposition 3.3. *Assume that $\sum_{j=1}^{n}\gamma_j < 2n$ and $\varepsilon = 1/n$. Then there exists a constant $c > 0$ such that:*

$$\lim_{t\to\infty} ||\mathbb{E}[\theta(t)] - \theta^*||_\infty \leq \frac{c}{n}. \tag{8}$$

Moreover, there exists a vector V and a constant c' such that:

$$\lim_{t\to\infty} \left\| \mathbb{E}[\theta(t)] - \theta^* - \frac{V}{n} \right\|_\infty \leq \frac{c}{n^2}. \tag{9}$$

Proof. When $\varepsilon = 1/n$, the evolution equation (2) can be rewritten (for all i and k) as:

$$\theta_i^k(t+1) = \varepsilon \sum_{j\in N} \mathbb{1}_{\{a_j(t)=k\}}.$$

This implies that for all $i, j \in N$, we have $\theta_i^k(t+1) = \theta_j^k(t+1)$.

This implies that the model falls in the category of model studied by [GLM18], where their variables $X_i(t)$ correspond to our variables $a_i(t)$, and their variables $M_k(t)$ is our $\theta_j^k(t)$ (for any j since they are all equal). Our result is a direct consequence of Corollary 1 of [GLM18]. □

Combining the ergodic theorem with Proposition 3.3 implies that

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \theta(t) = \theta^* + \frac{1}{n}V + O(\frac{1}{n^2}).$$

This justifies why the mean field approximation works so well: the expected opinions are at a distance of order $1/n$ from the approximation.

3.4. Homogeneous rationality

After exposing stability results of the fixed point in a general setting, we now focus on homogeneous population of agent, where all agents share the same rationality parameter $\gamma_j = \gamma$. As a particular case from the framework developed in Section 3.2, we obtain that fixed point uniqueness fails when the rationality parameter becomes sufficiently high, in particular when $\gamma > m$.

Theorem 5. *If $\gamma > m$, then G admits at least m distinct non-uniform consensus fixed points of the form $(x, y, \dots, y) \in \Delta^m$, where $x > 1/m$ and $y = (1 - x)/(m - 1) < 1/m$.*

Proof. Assume a non-uniform fixed point where one action is preferred and all others are symmetric.

$$c_1 = x > \frac{1}{m},$$
$$c_2 = c_3 = \cdots = c_m = y = \frac{1 - x}{m - 1},$$

with $x \neq y$ and $x + (m - 1)y = 1$.

To be a fixed point, it must satisfy the general fixed-point equation for the consensus vector c for each action. From the fixed point condition for action 1:

$$x = \frac{\exp(\gamma c_1)}{\sum_{\ell=1}^{m} \exp(\gamma c_\ell)} = \frac{\exp(\gamma x)}{\exp(\gamma x) + (m - 1)\exp(\gamma \frac{1-x}{m-1})}.$$

Similarly, for any other action we have:

$$y = \frac{\exp(\gamma y)}{\exp(\gamma x) + (m - 1)\exp(\gamma y)}.$$

Dividing these equations and taking logarithms gives:

$$\ln x - \ln y = \gamma(x - y). \tag{10}$$

Substituting $y = (1 - x)/(m - 1)$ into this equation, we define the function for $x \in (1/m, 1)$:

$$F(x) = \ln(x) - \ln\left(\frac{1 - x}{m - 1}\right) - \gamma\left(x - \frac{1 - x}{m - 1}\right).$$

At $x = 1/m$, we have $y = 1/m$ and:

$$F(1/m) = ln(1/m) - ln(\frac{1 - 1/m}{m - 1}) - \gamma(1/m - \frac{1 - 1/m}{m - 1}) = 0.$$

A value $x \in (1/m, 1)$ yields a fixed point of the form $(x, y, y, \cdots, y) \iff F(x) = 0$. The derivative is:

$$F'(x) = \frac{1}{x} + \frac{1}{1 - x} - \gamma(1 + \frac{1}{m - 1}) = \frac{1}{x(1 - x)} - \gamma\frac{m}{m - 1}$$

At $x = 1/m$, $F'(1/m) = \frac{1}{1/m(1-1/m)} - \gamma\frac{m}{m-1} = \frac{m(m-\gamma)}{m-1}$.

When $\gamma > m$, $F'(1/m) < 0$. For $x > 1/m$, $F(x) < 0$. Since $F(1/m) = 0$, there exists $\delta > 0$ such that for $x \in (1/m, 1/m + \delta)$, $F(x) < 0$.

As $x \to 1^-$, $\ln(y) \to -\infty \implies F(x) \to +\infty$. By the Intermediate Value Theorem, there $x^* \in (1/m, 1)$ such that $F(x^*) = 0$.

This yields a non-uniform fixed point $c^* = (x^*, y^*, \dots, y^*)$ with $y^* = (1 - x^*)/(m - 1)$. By symmetry of the action, there are m distinct fixed points (one for each action being the preferred one). This completes the proof.

□

We now provide a Bifurcation diagram example that shows the theoretical results established in this section.

Example 1. *As shown in Fig. 1, for $\gamma < m = 2$, numerical simulations from many dependent random initializations converge to the uniform consensus fixed point $c_{unif} = (0.5, 0.5)$, confirming Lemma 3.1 and Remark 1.*

At the critical value $\gamma = 2$, the system is at the switching point between the random and deterministic behavior. Small variations around c_{unif} appear, showing that the uniform state becomes unstable as predicted by Theorems 1 and 2, which establish that the consensus equilibrium loses stability when $\sum_{j=1}^{n} \gamma_j \geq 2n$.

For $\gamma > m = 2$, we observe two non-uniform fixed points across the random initializations, confirming Theorem 5. Each fixed point has the form $c^* = (\alpha, \beta)$ with $\alpha \neq \beta$. By symmetry, each action can become the dominant choice, leading to m distinct fixed points in the general case.

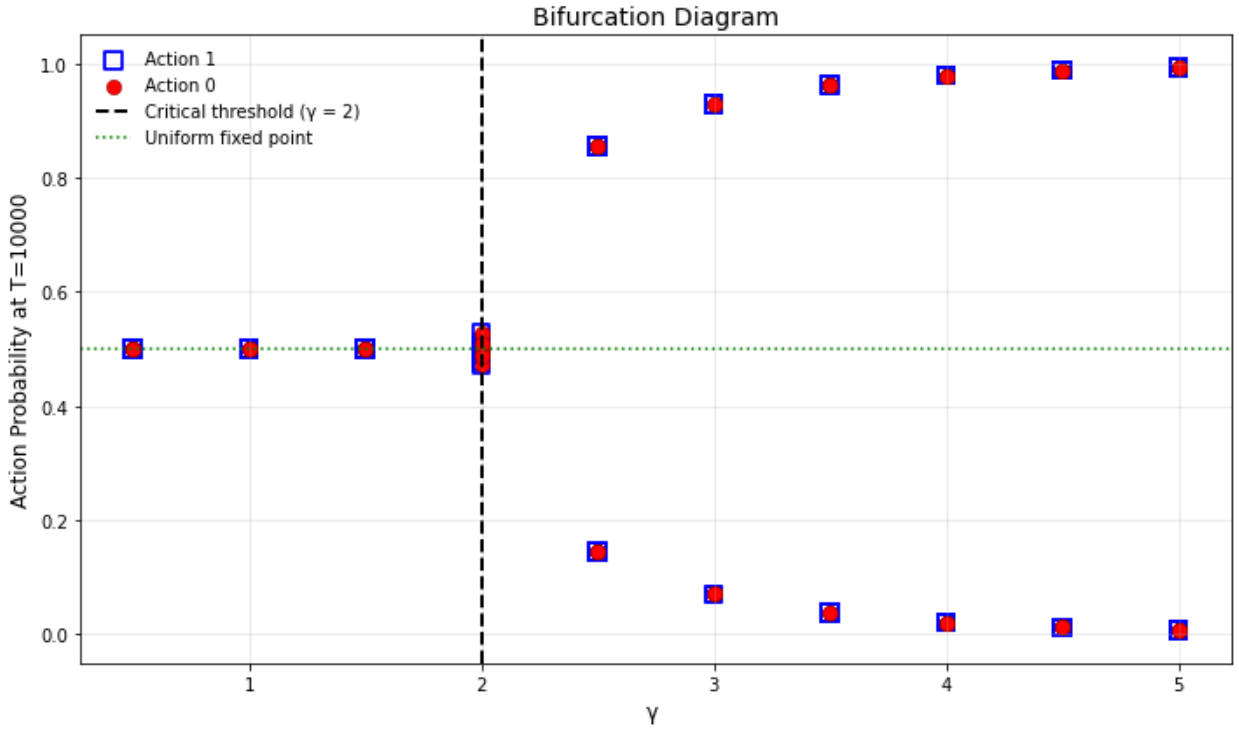

Figure 1: Bifurcation diagram for two actions ($m = 2$) system after $T = 10{,}000$ iterations. Each point corresponds to the final action probability obtained from random initialization. The vertical dashed line marks the critical threshold $\gamma = 2$, and the horizontal dotted line indicates the uniform fixed point $c_{\text{unif}} = (0.5, 0.5)$.

Remark 2. *Numerical simulations indicate that, in the homogeneous case, when $\gamma < m$, the uniform fixed point is the unique consensus equilibrium. For $\gamma > m$, we observe exactly m non uniform consensus fixed point of the symmetric form $(\alpha, \beta, \ldots, \beta)$, with $\alpha > 1/m > \beta$.*

The following section illustrates the behavior of the model through numerical experiments.

4. Simulations Results

First, we study how the rationality level affects agents' action distribution, then compare the stochastic dynamics with its deterministic mean-field under the same initial condition, and finally examine the model on a non-complete graph (connected/disconnected).

4.1. *Action distribution over time for all agents*

Fig. 2 shows, for each action $k \in \{1, \ldots, m\}$, the number of agents selecting k at each time step. We consider two groups: $N_1 = 50$ agents with high rationality ($\gamma_1 = 100$) and $N_2 = 50$ agents with low rationality ($\gamma_2 = 1$), with 10 actions.

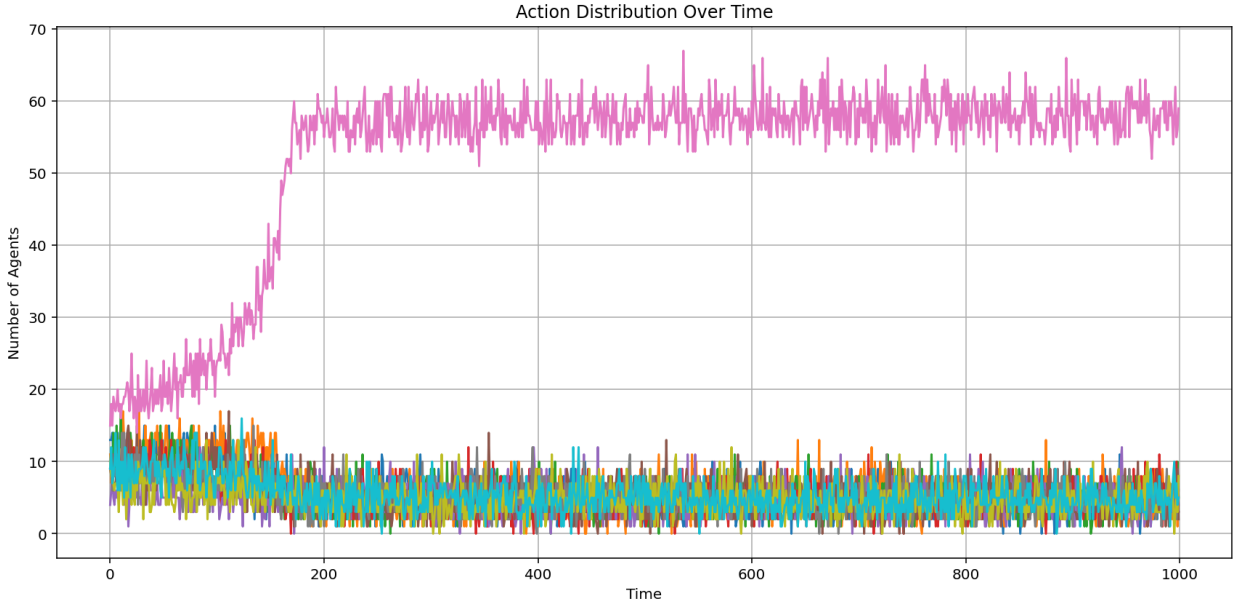

Figure 2: Action distribution over time (all agents). Each curve shows the number of agents choosing action k at time t. Parameters: $N_1 = N_2 = 50$, $m = 10$, $\gamma_1 = 100$, $\gamma_2 = 1$, $\varepsilon = 10^{-4}$, $T = 1000$.

The figure shows the appearance of a dominant action, here $k = 7$ (which corresponds to the pink curve), whose count around ≈ 58 agents. This behavior is explained by rational heterogeneity: the high-rationality group selects almost deterministically under the softmax rule, so once one action becomes preferred, it is consistently chosen by these agents. While low-rationality agents keep randomizing across actions. As a result, the dominant-action count is approximately N_1 plus a varying fraction of the N_2 low-rationality agents.

4.2. *Stochastic vs. Deterministic Approximation*

We compare the stochastic learning dynamics with its deterministic mean-field approximation to validate the theoretical model. For clarity, we focus on a simplified case: a homogeneous population where all agents share the same low rationality parameter $\gamma = 1$ and choose between $m = 2$ actions. The stochastic update uses the actual counts of how many agents chose each action ($\sum_j \mathbb{1}_{\{a_j(t)=k\}}$). The deterministic update uses the expected counts ($\sum_j \mathcal{F}_j^k(\theta_j(t))$) instead. Both simulations start with the same random initial preferences $\theta(0)$.

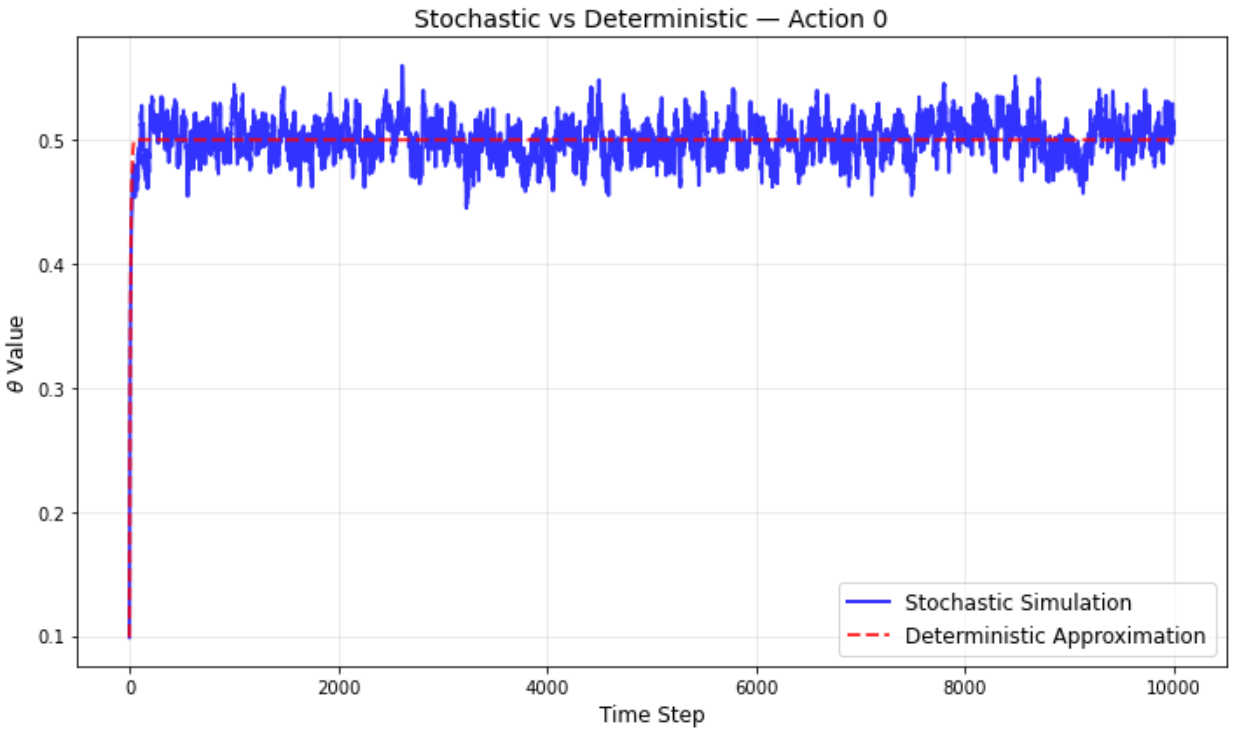

Figure 3: Preference for Action 0: stochastic vs. mean-field dynamics ($\gamma = 1, m = 2$).
The close match for a single agent validates the approximation G.

For the homogeneous case , the theoretical analysis predicts convergence to the uniform fixed point $(1/2, 1/2)$ since $\gamma < m$. Figure 3 shows the evolution of the preference for Action 0 for a representative agent. The stochastic trajectory (solid blue line) show small variations around the deterministic mean-field curves (dashed red line), demonstrating close alignment.

At the final time $T = 1000$, the stochastic simulation preferences $\theta_i(T) \approx [0.51, 0.49]$, while deterministic mean-field gives $\theta_i(T) \approx [0.5, 0.5]$. The mean error over time is 0.012 (approximately 1.5%). This close match confirms that the mean-field approximation accurately captures the average dynamics of the stochastic system for these parameter values. This example clearly illustrates the model's behavior and validates the theoretical prediction of uniform consensus when $\gamma < m$, providing a foundation for the more complex heterogeneous analyses.

4.3. Convergence of the Stochastic Dynamics as n Increases

Figure 4 illustrates the gap $\|\theta^{(n)}(t) - \theta^*\|_\infty$, where $\theta^{(n)}(t)$ is the stochastic state produced by the agent-level simulation (direct iteration of Eqs. 1 and 2) and θ^* is the mean-field fixed point. As n grows, the curve stays closer to 0, and its jumps are smaller. For larger n, the random system remains closer to the mean-field target.

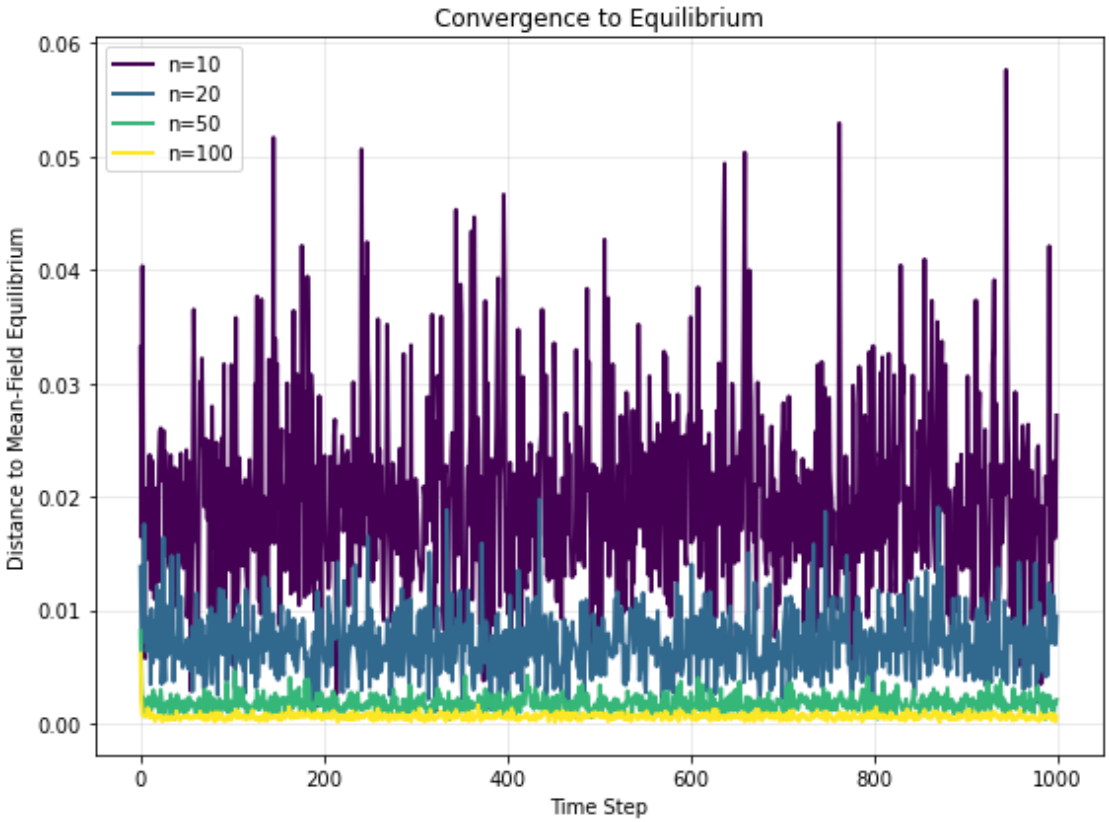

Figure 4: Distance to equilibrium for $n \in 10, 20, 50, 100$, where n is the number of agents.
Trajectories are generated by iterating Eqs. (1)–(2) (choosing actions via Eq. (1), then updating preferences via Eq. (2)). The error $\|\theta^{(n)}(t) - \theta^*\|_\infty$ decreases as n increases, illustrating concentration of the stochastic dynamics around the mean-field fixed point θ^*.

4.4. Learning on a non-complete graph

In a non-complete graph, each agent observes only the actions of its neighbors and applies the same two step procedure defined by Eqs. (1)–((2). We consider two networks, a connected sparse graph and a disconnected graph. In both cases, agents do not end with the same preferences, i.e., there exists $i \neq j : \theta_i(T) \neq \theta_j(T)$. Hence, the fixed points are not in consensus even when the network is connected. Figure 5 shows for both cases that the nodes are colored by the action that is preferred at time T ($\arg\max_k \theta_i^k(T)$). The presence of multiple colors confirms that the fixed point is not at consensus. This contrasts with the complete graph, where all actions are publicly observable, resulting in the existence of the consensus fixed points. Future work will characterize the conditions for the existence and stability of fixed points and their dependence on (ε, γ).

Remark 3. *We use a small network and three actions to provide a readable visualization of the interaction structure, while still reflecting the same update rules studied in the theory.*

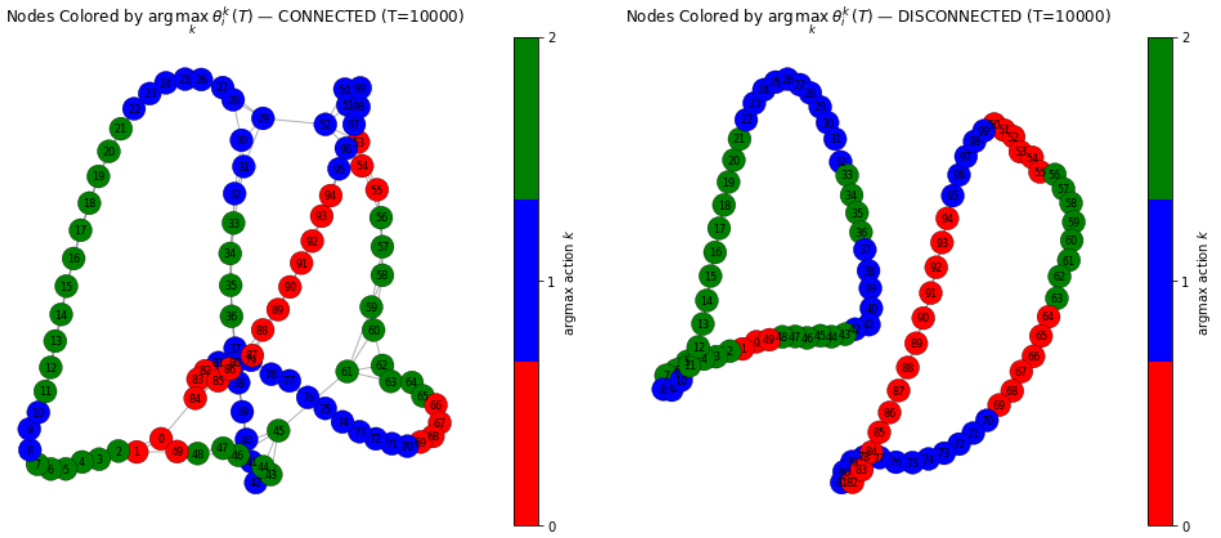

Figure 5: Agent-level simulation of the dynamics on connected and disconnected sparse graphs at time $T = 10^4$. Each node repeatedly chooses an action via Eq. (1) and updates its preference vector via Eq. (2) using only neighbor-observed actions. Nodes are colored according to their preferred action $\arg\max_k \theta_i^k(T)$. Parameters: $(N_1, N_2) = (50, 50)$, $K = 3$, $\gamma_1 = \gamma_2 = 50$ and $\varepsilon = 10^{-3}$. The presence of multiple colors in both networks shows that the dynamics do not reach consensus ($\exists\, i \neq j$: $\theta_i(T) \neq \theta_j(T)$).

5. Conclusion

In this work, we have analyzed a discrete-time Boltzmann social-learning system in which agents repeatedly choose actions while maintaining *private* preference and updating them from *public* observations of others' actions. Using a deterministic mean-field approximation map G, we proved the existence of fixed points and showed our results on a complete graphs. A central novelty of our work is that it explicitly accounts for *rational heterogeneity* by allowing agents to have distinct rationality parameters γ_i. This heterogeneity perspective enables us to characterize how the distribution of $\{\gamma_i\}$ shapes equilibrium selection and stability.

However, relies on some simplifying assumptions, as we consider a fixed action set (common m for all agents) and a non-dynamic population (fixed n over time). In addition, the main equilibrium structure results are derived for a complete graph model. This work opens several directions for future research. An immediate extension is to consider the dynamics on a network structure of agents, rather than a complete graph, and to study the structure of fixed points as well as their stability and convergence properties.

Acknowledgements

This work was supported by the French "Agence Nationale de la Recherche (ANR)", project UMICrowd under contract number ANR-22-CE38-0013.

References

[AG14] Özlem Ak Gümüş. Global and local stability analysis in a nonlinear discrete-time population model. *Advances in Difference Equations*, 11 2014.

[AG22] Sebastian Allmeier and Nicolas Gast. Mean field and refined mean field approximations for heterogeneous systems: It works! *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 6(1):1–43, 2022.

[BMS21] Virginia Bordignon, Vincenzo Matta, and Ali Sayed. Adaptive social learning. *IEEE Transactions on Information Theory*, PP:1–1, 07 2021.

[Cha04] Christophe Chamley. Rational herds: Economic models of social learning. *Rational Herds Economic Models of Social Learning*, 01 2004.

[CMMS16] N. Chowdhury, I. Morarescu, Samuel Martin, and Srikant Sukumar. Continuous opinions and discrete actions in social networks: a multi-agent system approach, 02 2016.

[Ela05] Saber Elaydi. An introduction to difference equation. 01 2005.

[FP11] Enrico Franchi and Agostino Poggi. *Multi-Agent Systems and Social Networks*. 01 2011.

[GHB24a] David Goll, Jobst Heitzig, and Wolfram Barfuss. Deterministic model of incremental multi-agent boltzmann q-learning: Transient cooperation, metastability, and oscillations, 2024.

[GHB24b] David Goll, Jobst Heitzig, and Wolfram Barfuss. Deterministic model of incremental multi-agent boltzmann q-learning: Transient cooperation, metastability, and oscillations, 12 2024.

[GLM18] Nicolas Gast, Diego Latella, and Mieke Massink. A refined mean field approximation of synchronous discrete-time population models. *Performance evaluation*, 126:1–21, 2018.

[Gol12] Russell Golman. Homogeneity bias in models of discrete choice with bounded rationality. *Journal of Economic Behavior Organization*, 82(1):1–11, 2012.

[GP17] Bolin Gao and Lacra Pavel. On the properties of the softmax function with application in game theory and reinforcement learning. 04 2017.

[HB24] Aamal Hussain and Francesco Belardinelli. Stability of multi-agent learning in competitive networks: delaying the onset of chaos. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'24/IAAI'24/EAAI'24. AAAI Press, 2024.

[HLL19] Shuyue Hu, Chin-wing Leung, and Ho-fung Leung. Modelling the dynamics of multiagent q-learning in repeated symmetric games: a mean field theoretic approach. *Advances in Neural Information Processing Systems*, 32, 2019.

[JG10] Matthew Jackson and Benjamin Golub. Naïve learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2:112–49, 02 2010.

[Lig05] Thomas M. Liggett. *Interacting Particle Systems*. Classics in Mathematics. Springer, Berlin, Heidelberg, 2005. Reprint of the 1985 original.

[LT21] Nadia Loy and Andrea Tosin. Boltzmann-type equations for multi-agent systems with label switching. *Kinetic and Related Models*, 14:867–894, 10 2021.

[LTV25] A. Leonidov, A. Titov, and E. Vasilyeva. Strategic exploitation in boltzmann q-learning in the prisoners dilemma. *Chaos, Solitons Fractals*, 201:117243, 12 2025.

[Mar07] Andre Martins. Continuous opinions and discrete actions in opinion dynamics problems. *International Journal of Modern Physics C*, 19, 12 2007.

[NMP17] Sarah Nowak, Luke Matthews, and Andrew Parker. A general agent-based model of social learning. *Rand health quarterly*, 7:10, 01 2017.

[PDPK15] NhatHai Phan, Dejing Dou, Brigitte Piniewski, and David Kil. Social restricted boltzmann machine: Human behavior prediction in health social networks. In *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 424–431, 2015.

[SEM05] Krzysztof Suchecki, Víctor Eguíluz, and Maxi Miguel. Voter model dynamics in complex networks: Role of dimensionality, disorder, and degree distribution. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 72:036132, 10 2005.

[SLA20] Brian Sifringer, Virginie Lurkin, and Alexandre Alahi. Enhancing discrete choice models with representation learning. *Transportation Research Part B: Methodological*, 140:236–261, 2020.

[SLB09] Yoav Shoham and Kevin Leyton-Brown. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, Cambridge, UK, 2009.

[TDBL25] Max Taylor-Davies, Neil Bramley, and Christopher Lucas. A rational framework for group-based selective social learning. *Open Mind*, 9:677–708, 05 2025.

[TNB12] Razieh Tadayon Nabavi and Mohammad Bijandi. Bandura's social learning theory social cognitive learning theory. 01 2012.

[WDK+25] Charley Wu, Dominik Deffner, Benjamin Kahl, Björn Meder, Mark Ho, and Ralf Kurvers. Adaptive mechanisms of social and asocial learning in immersive collective foraging. *Nature Communications*, 16, 04 2025.

[YLL+18] Yaodong Yang, Rui Luo, Minne Li, Ming Zhou, Weinan Zhang, and Jun Wang. Mean field multi-agent reinforcement learning. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 5571–5580. PMLR, 10–15 Jul 2018.