

Neumann, Berenice Anne

Article — Published Version

A myopic adjustment process for mean field games with finite state and action space

International Journal of Game Theory

Provided in Cooperation with:

Springer Nature

Suggested Citation: Neumann, Berenice Anne (2023) : A myopic adjustment process for mean field games with finite state and action space, International Journal of Game Theory, ISSN 1432-1270, Springer, Berlin, Heidelberg, Vol. 53, Iss. 1, pp. 159-195,
<https://doi.org/10.1007/s00182-023-00866-z>

This Version is available at:

<https://hdl.handle.net/10419/307524>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



A myopic adjustment process for mean field games with finite state and action space

Berenice Anne Neumann¹ 

Accepted: 22 June 2023 / Published online: 1 August 2023
© The Author(s) 2023

Abstract

In this paper, we introduce a natural learning rule for mean field games with finite state and action space, the so-called myopic adjustment process. The main motivation for these considerations is the complexity of the computations necessary to determine dynamic mean field equilibria, which makes it seem questionable whether agents are indeed able to play these equilibria. We prove that the myopic adjustment process converges locally towards strict stationary equilibria under rather broad conditions. Moreover, we also obtain a global convergence result under stronger, yet intuitive conditions.

Keywords Mean field games · Learning in games · Finite state space · Finite action space

JEL Classification C73 · C70

1 Introduction

Mean field games have been introduced by Lasry and Lions (2007) and Huang et al. (2006) in order to make dynamic games with a large number of players tractable. The central idea is to approximate these games with many players by a game with a continuum of anonymous players. Thereafter a vibrant field of research emerged in particular concerning games where the dynamics of the individual players are described by diffusions. For a first overview consider the monographs of Bensoussan et al. (2013) and Carmona and Delarue (2018a, 2018b) or the lecture notes by Cardaliaguet (2013). However, a central problem for applications is that equilibria are described by forward-backward systems of (stochastic) differential equations and are, therefore, notoriously intractable.

✉ Berenice Anne Neumann
neumannb@uni-trier.de

¹ Trier University, Universitätsring 19, 54296 Trier, Germany

Recently, also mean field games with finite state space have been considered, for example in Gomes et al. (2013), Cecchin and Fischer (2018), Belak et al. (2021), Carmona and Wang (2021), Doncel et al. (2019), Neumann (2020) and Carmona and Delarue (2018a, Section 7.2). Also in these games dynamic equilibria are described by forward-backward systems of differential equations. Moreover, considering stationary equilibria allows for closed-form solutions: This was first demonstrated in several applications including the spread of corruption, botnet defence, paradigm shift in science and consumer choice (Besancenot and Dogguy 2015; Kolokoltsov and Bensoussan 2016; Kolokoltsov and Malafeyev 2017; Gomes et al. 2014). Thereafter, in Neumann (2020) results yielding a semi-explicit characterization of stationary equilibria in general mean field games with finite state and action space have been derived.

For applications it is now clearly desirable to understand whether stationary equilibria are an adequate description of agents' behaviour. One approach is to discuss the question in how far these equilibria are suitable limit objects of dynamic equilibria when the finite time horizon tends to infinity (consider Kolokoltsov and Malafeyev (2018) for the analysis of an example). The second approach is to understand in how far stationary equilibria arise when agents apply certain only partly rational decision rules. This approach is a classical one in standard game theory and known as learning.

Why players should play equilibrium strategies is a classical concern. The explanation that agents arrive at these strategies from "introspection and computation" is challenged by many facts: The computational complexity of the problems at hand, the question which equilibrium to choose in case of multiple equilibria as well as experimental evidence (Fudenberg and Levine 1998). Because it was observed in experiments that agents "learn" to play equilibria after some time, many authors focussed on the definition and analysis of partially rational "learning rules" (for example fictitious play or partial best response) mostly for static games. For an overview consider the monograph of Fudenberg and Levine (1998) or the survey by Nachbar (2009).

Recently, learning has also been discussed for mean field games with diffusion-based dynamics: On the one hand, in Cardaliaguet and Hadikhanloo (2017) fictitious play for repeated games with finite time horizon has been introduced, which thereafter has been analysed in subsequent publications (Hadikhanloo 2017; Briani and Cardaliaguet 2018) and extended to discrete time finite state mean field games with finite time horizon (Hadikhanloo 2018). On the other hand, in Mouzouni (2018) a learning procedure similar to the myopic adjustment process considered here has been introduced. In this paper existence and local convergence under strong assumptions (quadratic Hamiltonian and Lasry–Lions monotonicity condition) have been proved. We highlight, that the methods for diffusion-based mean field games cannot be adapted to our setting of a mean field game with finite state and action space, since the crucial assumption for most methods (a unique optimizer of the Hamiltonian) is typically not satisfied (see Neumann 2020, Remark 2.5).

Another branch of literature analyses learning methods for discrete time mean field games with finite state and action space from a reinforcement learning perspective, see the survey by Laurière et al. (2022) for an overview. These learning

methods do not aim to explain how players adjust their behaviour in a (partially) rational way, but how to obtain nearly optimal solutions using data or samples. Hence, these methods include (generalizations of) classical (economically motivated) learning rules like fictitious play (Perrin et al. 2020) or best-response based methods (Guo et al. 2019), but also others.

Up to the knowledge of the author, this paper is the first paper that considers learning in continuous time mean field games with finite state and action space. The learning procedure we consider in this paper is motivated from an economic standpoint as a partially rational decision procedure for the agents. Namely, we introduce a myopic adjustment process, where agents choose to play a best response for the scenario that the current state distribution in the population will persist for all future times. Moreover, whenever the current state distribution in the population changes, the agent will adjust his strategy to the new best response. We highlight that, in contrast to most other learning procedures, we do not assume that all agents choose the same strategy as best response. Given the players' choices the state distribution of the population gradually changes as described by the chosen strategies. This differs from best-response based methods, where the distribution would immediately change to the stationary distribution given the best response. This definition yields to a formulation of this process as a differential inclusion and we can prove existence of this process under a continuity assumption. Thereafter we address the question whether the process converges locally or globally towards stationary equilibria. In this context we first obtain under suitable conditions an analogue of the classical result from the theory of matrix games that a strict equilibrium is locally stable. Thereafter, we also establish a global convergence result under stronger, yet intuitive conditions, which are different to the conditions used in analogous contexts in evolutionary game theory. Since the proof is rather complex and technical we first provide a result for a two strategy setting in which case the general idea becomes clear. Thereafter, we provide the general statement.

Let us conclude the introduction by relating the myopic adjustment process with the notion of evolutionary game dynamics for population games as in Sandholm (2010, 2015). In the setting of a static game among a continuum of small, anonymous players, an evolutionary game dynamic describes the evolution of the strategy choices using so-called revision protocols, which are simple myopic rules to adjust the behaviour based on the observed strategies by all others. The myopic adjustment process considered here can be seen as an extension of this idea to dynamic games where each agent has an individual state that evolves over time (according to the action he chooses) and earns a reward that depends on his individual state, his individual action and the distribution of states of the other agents. Thus, in contrast to standard learning procedures the game is not played repeatedly and the agents do not learn from round to round, but the agents learn while playing the game. A second difference is that besides the strategies also a state evolves over time, an idea that occurs also in Marden (2012). In this paper a global state taking finitely many values that evolves over time is introduced and the static game that is played repeatedly depends on this state. However, here we introduce an individual state for each agent, which has, up to the knowledge of the author, not been considered so far. Due to fact that the adjustment of strategies and the evolution of the individual states happen

simultaneously, classical results from evolutionary game theory cannot be applied. However, since the myopic adjustment process looks somehow similar to classical evolutionary game dynamics the results often have the same flavour. In this sense, we prove as in Sandholm (2014) that the myopic adjustment process converges locally towards strict equilibria and similar to Hofbauer and Sandholm (2009) and Zusai (2020) we propose gradient conditions to ensure global convergence of the adjustment process.

The rest of the paper is structured as follows: Sect. 2 describes the mean field game model considered in this paper. Section 3 introduces the myopic adjustment process and justifies its definition as a sensible partially rational learning rule. Moreover, it presents the myopic adjustment process for a simple example. In Sect. 4 we study the local convergence of the myopic adjustment process and in Sect. 5 we investigate the global convergence first for the special case of two strategies, thereafter in a general setting. The Appendix A describes an algorithm to verify the conditions of the general global convergence theorem, Appendix B contains all proofs.

2 Stationary equilibria of mean field games with finite state and action space

This section describes the mean field game model. The setup is the same as in Neumann (2020) and we refer the reader to this paper for more details. Moreover, we remark that the model has been first introduced in an analytic formulation and without the notion of stationary equilibria in Doncel et al. (2019).

We consider a continuum of agents. Let $S = \{1, \dots, S\}$ ($S > 1$) be the set of possible states of each player and let $\mathcal{A} = \{1, \dots, A\}$ be the set of possible actions. With $\mathcal{P}(S)$ we denote the probability simplex over S and with $\mathcal{P}(\mathcal{A})$ the probability simplex over \mathcal{A} . We refer to an element $m \in \mathcal{P}(S)$ that describes the distribution of states in the population as *social state*, whereas the state $i \in S$ of an individual agent is called the individual state. A (*mixed*) *strategy* is a measurable function $\pi : S \times [0, \infty) \rightarrow \mathcal{P}(\mathcal{A})$, $(i, t) \mapsto (\pi_{ia}(t))_{a \in \mathcal{A}}$ with the interpretation that $\pi_{ia}(t)$ is the probability that at time t and in the individual state i the player chooses action a . A strategy $\pi = d : S \times [0, \infty) \rightarrow \mathcal{P}(\mathcal{A})$ is *deterministic* (or a *pure strategy*) if it satisfies for all $t \geq 0$ and for all $i \in S$ that there is an $a \in \mathcal{A}$ such that $d_{ia}(t) = 1$ and $d_{ia'}(t) = 0$ for all $a' \in \mathcal{A} \setminus \{a\}$. Sometimes the following equivalent representation is helpful: Namely, we represent a deterministic strategy as a function $d : S \times [0, \infty) \rightarrow \mathcal{A}$, $(i, t) \mapsto d_i(t)$ with the interpretation that $d_i(t) = a$ states that at time t in the individual state i action a is chosen. A *stationary strategy* is a map $\pi : S \times [0, \infty) \rightarrow \mathcal{P}(\mathcal{A})$ such that $\pi_{ia}(t) = \pi_{ia}$ for all $t \geq 0$. With Π we denote the set of all (mixed) strategies and with Π^s the set of all stationary strategies. Similarly, we denote by D the set of all deterministic strategies and by D^s the set of all deterministic stationary strategies.

Let for all $a \in \mathcal{A}$ and $m \in \mathcal{P}(S)$ the matrices $(Q_{\cdot\cdot a}(m))_{a \in \mathcal{A}}$ be conservative generators, that is $Q_{ija}(m) \geq 0$ for all $i, j \in S$ with $i \neq j$ and $\sum_{j \in S} Q_{ija}(m) = 0$ for all $i \in S$. The individual dynamics of each player given a Lipschitz continuous

flow of social states $m : [0, \infty) \rightarrow \mathcal{P}(\mathcal{S})$ and a strategy $\pi : \mathcal{S} \times [0, \infty) \rightarrow \mathcal{P}(\mathcal{A})$ are given as a Markov process $X^\pi(m)$ with given initial distribution $x_0 \in \mathcal{P}(\mathcal{S})$ and infinitesimal generator given by the $Q(t)$ -matrix

$$(Q^\pi(m(t), t))_{ij} = \sum_{a \in \mathcal{A}} Q_{ija}(m(t)) \pi_{ia}(t).$$

Given the initial condition $x_0 \in \mathcal{P}(\mathcal{S})$, the goal of each player is to maximize his expected discounted reward, which is given by

$$V_{x_0}(\pi, m) = \mathbb{E} \left[\int_0^\infty \left(\sum_{a \in \mathcal{A}} r_{X^\pi(m)a}(m(t)) \pi_{X^\pi(m)a}(t) \right) e^{-\beta t} dt \right], \quad (1)$$

where $r : \mathcal{S} \times \mathcal{A} \times \mathcal{P}(\mathcal{S}) \rightarrow \mathbb{R}$ is a real-valued function and $\beta \in (0, 1)$ is the discount factor. That is, for a fixed flow of social states $m : [0, \infty) \rightarrow \mathcal{P}(\mathcal{S})$ the individual agent's decision problem is a Markov decision process with expected discounted reward criterion and time-inhomogeneous reward functions and transition rates.

In this paper we work under the following standing assumption, which ensures the well-definition of the model as well as the existence of dynamic as well as stationary equilibria (Neumann 2020):

Assumption A1 For all $i, j \in \mathcal{S}$ and all $a \in \mathcal{A}$ the function $m \mapsto Q_{ija}(m)$ mapping from $\mathcal{P}(\mathcal{S})$ to \mathbb{R} is Lipschitz-continuous in m . For all $i \in \mathcal{S}$ and all $a \in \mathcal{A}$ the function $m \mapsto r_{ia}(m)$ mapping from $\mathcal{P}(\mathcal{S})$ to \mathbb{R} is continuous in m .

Definition 2.1 Given an initial distribution $m_0 \in \mathcal{P}(\mathcal{S})$, a *mean field equilibrium* is a pair (m, π) consisting of a flow of social states $m : [0, \infty) \rightarrow \mathcal{P}(\mathcal{S})$ with $m(0) = m_0$ and a strategy $\pi : \mathcal{S} \times [0, \infty) \rightarrow \mathcal{P}(\mathcal{A})$ such that

- the distribution of the process $X^\pi(m)$ at time t is given by $m(t)$, and
- $V_{m_0}(\pi, m) \geq V_{m_0}(\pi', m)$ for all $\pi' \in \Pi$.

Definition 2.2 A stationary mean field equilibrium is given by a stationary strategy π and a vector $m \in \mathcal{P}(\mathcal{S})$ such that

- the law of $X^\pi(m)$ at any point in time t is given by m , and
- for any initial distribution $x_0 \in \mathcal{P}(\mathcal{S})$ we have $V_{x_0}(\pi, m) \geq V_{x_0}(\pi', m)$ for all $\pi' \in \Pi$.

As discussed in Neumann (2020) this is a sensible notion of stationary equilibrium: Indeed, the second condition ensures that an agent will at no time benefit from deviating from the equilibrium strategy since irrespective of his current state or distribution, respectively, the strategy π is optimal for him.

3 The myopic adjustment process

In general, it is not possible to compute dynamic mean field equilibria for the considered game; it is not even possible to explicitly characterize solutions of the individual control problem for a given non-constant flow of social states. Moreover, also in the case of a finite time horizon, the search for equilibria can only be reduced to a forward-backward system of ODEs, which can, most of the time, be only solved numerically (see Belak et al. 2021). The aim of this section is to motivate and define a reasonable alternative decision mechanism for the agents.

In contrast to Cardaliaguet and Hadikhanloo (2017) we cannot assume that the game is played repeatedly, but instead we have to assume that the agent changes his strategy during the game. In contrast to classical evolutionary game theory, we moreover assume that the agents can change their strategy at any time t . We note that the game at time t with current distribution m is, due to the time-homogeneous formulation and the infinite time horizon, equivalent to the game started at time 0 with initial distribution m . Moreover, we remind ourselves that the influence of the individual agent on the game characteristics and thus on the payoff of the other players is negligible. Therefore, it is reasonable to assume that the agents do not try to influence the other players' choices, but that they only maximize their own payoff. Because of time-homogeneity and the negligible influence on other players, we assume that the agents choose Markovian strategies that only depend on the current individual state and current social state.

We assume that the agent when choosing an optimal strategy given the current social state m assumes that the social state is constant. Indeed, by construction of the dynamics and since Q is uniformly bounded, the social state will be close to m for a certain time horizon. Since the rewards and transition rates are continuous, this means that for this time horizon the approximation that the social state is constant works well. For a longer time horizon a sensible prediction of the evolution of the social state is complex, since the agent would also need to take into account that the population's strategy might change due to the change in the social state. However, since the agent is allowed to change his strategy at any time (in particular if the social state changes drastically), it is reasonable to assume that he focuses on the near future, i.e. chooses a strategy that maximizes his reward given the constant prediction m .

Given such a constant prediction of the evolution of the social state, the optimization problem becomes a tractable Markov decision process with stationary transition rates and rewards (see Neumann 2020, Lemma 3.1). It is well known that there is always an optimal stationary strategy for the considered optimization problem (Guo and Hernández-Lerma 2009) and it is again natural to assume that agents choose such a stationary strategy. We remark that this assumption that agents choose a stationary strategy is classical and that there are several conceptual reasons for the use of these strategies (see Maskin and Tirole 2001). Namely, stationary Markov strategies are the simplest (rational) form of decision-making in this context. Moreover, this type of strategies is related to subgame perfection, which is as discussed earlier a reasonable requirement in our setting. Indeed, these strategies ensure that a game with the same relevant characteristics (i.e. individual and social state) is played in

the same way. Finally, the restriction on this type of strategies reduces the number of possible best responses and thus increases predictive power.

In order to characterize the optimal stationary strategies that could be chosen by the agent let us review the relevant results on Markov decision processes with stationary transition rates and rewards. By $V^\pi(m) = (V_{\delta_i}^\pi(m))_{i \in \mathcal{S}}$ we denote the reward vector for the Markov decision process, which collects for each $i \in \mathcal{S}$ the expected discounted reward defined in (1) given that the individual agent starts in state i , chooses strategy $\pi \in \Pi$ and the social state is m for all times. A strategy π is optimal if it satisfies $V^\pi(m) \geq V^{\tilde{\pi}}(m)$ pointwise for all $\tilde{\pi} \in \Pi$. Moreover, if π is optimal we have $V^\pi(m) = V^*(m)$, where $V^*(m)$ is the unique solution of the optimality equation

$$\beta V_i^*(m) = \max_{a \in \mathcal{A}} \left\{ r_{ia}(m) + \sum_{j \in \mathcal{S}} Q_{ija}(m) V_j^*(m) \right\}, \quad i \in \mathcal{S}.$$

For our purpose now the following result explicitly characterizing the set of all optimal stationary strategies proves to be useful (see Neumann 2020, Section 3): Define

$$O_i(m) =: \operatorname{argmax}_{a \in \mathcal{A}} \left\{ r_{ia}(m) + \sum_{j \in \mathcal{S}} Q_{ija}(m) V_j^*(m) \right\}$$

and set

$$\mathcal{D}(m) := \{d : \mathcal{S} \rightarrow \mathcal{A} \mid d(i) \in O_i(m) \text{ for all } i \in \mathcal{S}\}.$$

Then the set of all optimal stationary strategies is given by $\operatorname{conv}(\mathcal{D}(m))$.

So all in all we assumed that at any time any agent can change his strategy and he will do this using his current individual state and the current social state. Moreover, we argued that every agent will choose a strategy from the set $\operatorname{conv}(\mathcal{D}(m))$. However, we cannot assume that all agents choose a particular strategy nor that the agents or groups of them agree on a common strategy. Moreover, we also cannot describe which agents will choose which strategy. The only sensible assumption is that the population chooses aggregately a strategy from the set $\operatorname{conv}(\mathcal{D}(m))$. The next lemma describes how the social state evolves if all agents adopt this decision mechanism:

Lemma 3.1 *Let $m : [0, \infty) \rightarrow \mathcal{P}(\mathcal{S})$ be the distribution of the population, where at time $t \geq 0$ any agent chooses a strategy from $\operatorname{conv}(\mathcal{D}(m(t)))$. Then*

$$\dot{m}(t) \in F(m(t)) := \operatorname{conv} \left\{ (Q^d(m(t)))^T m(t) : d \in \mathcal{D}(m(t)) \right\} \quad (2)$$

for almost all $t \geq 0$.

With these preparations we define the myopic adjustment process as a solution of the differential inclusion in the sense of Deimling (1992) given by (2). Namely, a trajectory of the myopic adjustment process is an absolutely continuous function $m : [0, \infty) \rightarrow \mathcal{P}(\mathcal{S})$ such that

$$\dot{m}(t) \in F(m(t)) \quad \text{for almost all } t \geq 0, \quad m(0) = m_0. \quad (3)$$

We remark, that the use of differential inclusions as a modelling tool for situations where uncertainty, the absence of control or a variety of available dynamics occurs is classical (Aubin and Cellina 1984).

Before we start the analysis of the long-term behaviour, we note, relying on a classical existence result for differential inclusions, that under Assumption A1 a solution of the differential inclusion exists.

Theorem 3.2 *The differential inclusion defined by (2) and (3) admits a solution $m : [0, \infty) \rightarrow \mathcal{P}(S)$.*

Let us briefly compare the myopic adjustment process to classical evolutionary game dynamics. The central difference is that here two layers of adjustment have to be considered, namely the adjustment of strategies and the evolution of the individual states, whereas in classical evolutionary game dynamics only the adjustment of the strategies matters. Also from the motivation/construction the processes differ: First, here we consider learning during a dynamic game, whereas in classical evolutionary game dynamics it is assumed that agents learn from round to round in a static game. Second, here the agent is allowed to change his strategy at any time, in classical evolutionary game dynamics he can often do this only at random times. This means that in evolutionary game dynamics the change of strategies is rather continuous, whereas here it is abrupt and sometimes discontinuous. However, since the individual agent's states are modelled by a Markov chain and the map $m \mapsto \mathcal{D}(m)$ is upper semi-continuous (see the proof of Theorem 3.2), the resulting evolution of the social state is again upper semi-continuous. Yet, besides all the differences, the emerging processes are both dynamical systems on a probability simplex having a similar structure. Therefore, also the results have a similar flavour, although the economic intuitions behind the results are different.

For our purpose, it is central to understand how stationary equilibria and the trajectories of the myopic adjustment process interact. The following observation, which is classical for many learning procedures, is a first step:

Remark 3.3 By definition, a point is a stationary point of (3) if and only if it is a stationary mean field equilibrium. This is immediate, since $0 \in F(m(t))$ implies that

$$\begin{aligned} 0 &\in \operatorname{conv} \left\{ \left(\sum_{i \in S} \sum_{a \in A} m_i Q_{ija}(m(t)) d_{ia} \right)_{j \in S} : d \in \mathcal{D}(m(t)) \right\} \\ &= \left\{ \left(\sum_{i \in S} \sum_{a \in A} m_i Q_{ija}(m(t)) \pi_{ia} \right)_{j \in S} : \pi \in \operatorname{conv}(\mathcal{D}(m(t))) \right\}. \end{aligned}$$

□

In Sect. 4 we will analyse, whether the myopic adjustment process started close to a stationary equilibrium converges towards it. Thereafter, in Sect. 5 we analyse under which conditions convergence towards some stationary equilibrium irrespective of the starting point can be expected.

We conclude this section by discussing the shape of the myopic adjustment process for an example linked to consumer choice in the mobile phone sector, for which previously in Neumann (2020) the stationary equilibria have been computed:

Example 3.4 The agents can choose between two providers and their utility is increasing in the share of customers using the same provider. The agents can switch the provider facing a time-unit cost c . However, the decision is not implemented immediately, but according to a Poisson process with rate b . The formal description of the model is given as follows: Let $\mathcal{S} = \{1, 2\}$ and $\mathcal{A} = \{stay, change\}$. Let $\delta > 0$ be small and define

$$f_\delta : \mathbb{R} \rightarrow \mathbb{R} \quad y \mapsto \begin{cases} \frac{1}{2\delta}y^2 + \frac{\delta}{2} & \text{if } y \leq \delta \\ y & \text{if } y > \delta \end{cases}.$$

Then the transition rates and rewards are given by

$$\begin{aligned} Q^{change}(m) &= \begin{pmatrix} -b & b \\ b & -b \end{pmatrix} & r^{change}(m) &= \begin{pmatrix} \ln(f_\delta(m_1)) + s_1 - c \\ \ln(f_\delta(1 - m_1)) + s_2 - c \end{pmatrix} \\ Q^{stay}(m) &= \begin{pmatrix} -\epsilon & \epsilon \\ \epsilon & -\epsilon \end{pmatrix} & r^{stay}(m) &= \begin{pmatrix} \ln(f_\delta(m_1)) + s_1 \\ \ln(f_\delta(1 - m_1)) + s_2 \end{pmatrix}, \end{aligned}$$

where ϵ , b , s_1 , s_2 and c are positive constants with $\epsilon < b$. Note that $m_2 = 1 - m_1$, however, we need this unusual definition for the analysis of the global convergence.

Let us introduce the notation cs for the strategy d such that $d(1) = change$ and $d(2) = stay$ and analogously the notations sc and ss . In Neumann (2020) it is then shown that

$$\mathcal{D}(m) = \begin{cases} \{cs\} & \text{if } m_1 < k_1 \\ \text{conv}\{cs, ss\} & \text{if } m_1 = k_1 \\ \{ss\} & \text{if } k_1 < m_1 < k_2, \\ \text{conv}\{ss, sc\} & \text{if } m_1 = k_2 \\ \{sc\} & \text{if } m_1 > k_2 \end{cases}$$

where

$$k_1 = \frac{\exp\left(-\frac{c(\beta+2\epsilon)}{b-\epsilon} - s_1 + s_2\right)}{1 + \exp\left(-\frac{c(\beta+2\epsilon)}{b-\epsilon} - s_1 + s_2\right)} \quad \text{and} \quad k_2 = \frac{\exp\left(\frac{c(\beta+2\epsilon)}{b-\epsilon} - s_1 + s_2\right)}{1 + \exp\left(\frac{c(\beta+2\epsilon)}{b-\epsilon} - s_1 + s_2\right)}.$$

Thus, the myopic adjustment process is defined as a solution of the differential inclusion $\dot{m}(t) \in F(m(t))$ with

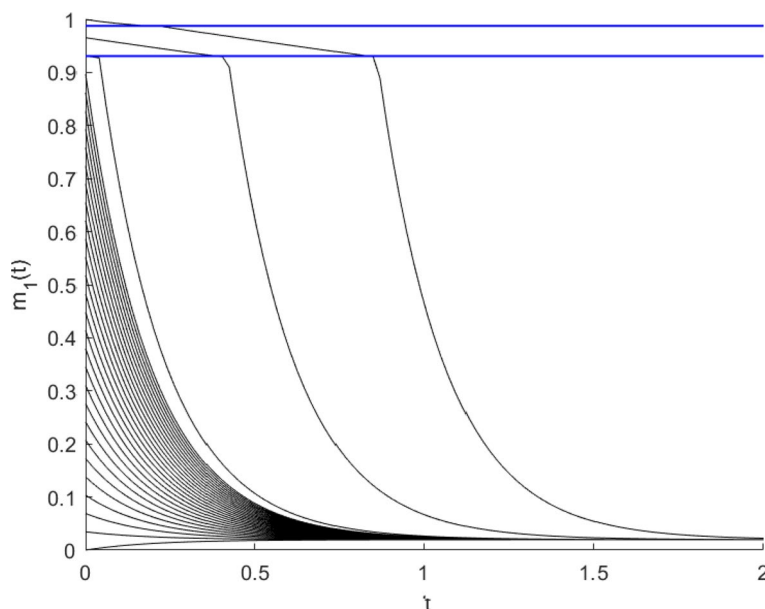


Fig. 1 Illustration of Example 3.4: The figure shows $m_1(t)$ for several trajectories of the myopic adjustment process given different initial conditions. The blue vertical lines represent the thresholds at which the set of optimal strategies changes

$$F(m) = \begin{cases} \begin{pmatrix} -bm_1(t) + \epsilon m_2(t) \\ bm_1(t) - \epsilon m_2(t) \end{pmatrix} & \text{if } m_1(t) < k_1 \\ \text{conv} \left\{ \begin{pmatrix} -bm_1(t) + \epsilon m_2(t) \\ bm_1(t) - \epsilon m_2(t) \end{pmatrix}, \begin{pmatrix} -\epsilon m_1(t) + \epsilon m_2(t) \\ \epsilon m_1(t) - \epsilon m_2(t) \end{pmatrix} \right\} & \text{if } m_1(t) = k_1 \\ \begin{pmatrix} -\epsilon m_1(t) + \epsilon m_2(t) \\ \epsilon m_1(t) - \epsilon m_2(t) \end{pmatrix} & \text{if } k_1 < m_1(t) < k_2 \\ \text{conv} \left\{ \begin{pmatrix} -\epsilon m_1(t) + \epsilon m_2(t) \\ \epsilon m_1(t) - \epsilon m_2(t) \end{pmatrix}, \begin{pmatrix} -\epsilon m_1(t) + bm_2(t) \\ \epsilon m_1(t) - bm_2(t) \end{pmatrix} \right\} & \text{if } m_1(t) = k_2 \\ \begin{pmatrix} -\epsilon m_1(t) + bm_2(t) \\ \epsilon m_1(t) - bm_2(t) \end{pmatrix} & \text{if } m_1(t) > k_2 \end{cases}$$

In Fig. 1 we illustrate the behaviour of this process for a parameter choice that yields for any initial condition to a unique solution of the differential inclusion.

4 Local convergence

This section discusses the question of local convergence, that is we analyse under which conditions we can expect that for an initial condition close to a stationary equilibrium the trajectories of (3) converge towards that equilibrium. We start the analysis of the case where the equilibrium is strict in the sense that the equilibrium strategy is the unique optimal strategy for the equilibrium social state.

Thereafter, we comment on the (limited) possibilities to extend the results to non-strict equilibria.

Lemma 4.1 *Let $\bar{m} \in \mathcal{P}(\mathcal{S})$ and assume that d is the unique optimal stationary strategy for \bar{m} (that is $\mathcal{D}(\bar{m}) = \{d\}$). Then there is an $\epsilon > 0$ such that for all $m' \in N_\epsilon(\bar{m})$ we have $\mathcal{D}(m') = \{d\}$.*

This lemma now yields that for all $m' \in N_\epsilon(\bar{m})$ we have $F(m') = \{(Q^d(m'))^T m'\}$. Thus, it suffices to investigate whether there is a $\delta > 0$ such that for all $m_0 \in \mathcal{P}(\mathcal{S})$ satisfying $|m_0 - \bar{m}| < \delta$ we have that the solution of $\dot{m}(t) = (Q^d(m(t)))^T m(t)$ lies in $N_\epsilon(\bar{m})$ for all $t \geq 0$ and converges towards \bar{m} . This question is closely linked to the notion of asymptotically stable solutions of autonomous ordinary differential equations $\dot{x} = f(x)$. However, we do not consider arbitrary initial conditions in $N_\delta(\bar{x})$, but only those that lie in $\mathcal{P}(\mathcal{S})$.

The first positive result we present covers the case where the dynamics given the equilibrium strategy are constant, that is $\dot{m} = (Q^d)^T m$.

Theorem 4.2 *Let (\bar{m}, d) be a stationary mean field equilibrium such that $\mathcal{D}(\bar{m}) = \{d\}$ (that is, d is the unique optimal strategy at \bar{m}). Furthermore, assume that $Q^d(m)$ is constant in m and an irreducible generator. Then there is a $\delta > 0$ such that any solution of the myopic adjustment process (3) with initial condition $m_0 \in \mathcal{P}(\mathcal{S}) \cap N_\delta(\bar{m})$ converges exponentially fast to \bar{m} , i.e. there are constants $C_1, C_2 > 0$ such that*

$$||m(t) - \bar{m}|| \leq C_1 e^{-C_2 t} \quad \text{for all } t \geq 0.$$

The condition that Q is an irreducible generator is a classical one in the analysis of Markov chains, it basically states that an agent being in any individual state has a positive probability to move to any other individual state. In this sense, the assumption requires that the state space is not partitioned into several parts that the individual agent cannot leave. It is immediately clear, that if the condition is not met, local convergence cannot be expected for all starting points. Indeed, if the share of agents in each of the parts of the state space does not coincide with the share of agents in the respective parts for the equilibrium social state, then convergence is not possible.

The proof of the statement now mainly relies on the fact that the eigenvalues of an irreducible generator are such that there is an eigenvalue $\lambda_1 = 0$ with multiplicity 1 and all other eigenvalues $\lambda_2, \dots, \lambda_n$ have negative real part (Asmussen 2003, Corollary 4.9). This insight together with the fact that the solutions of a linear ODE with initial condition $v \in \text{Eig}(\lambda)$ for some eigenvalue λ is given by $x(t) = e^{\lambda t} \sum_{l=0}^{m(\lambda)-1} t^l / l! ((Q^d)^T - \lambda I)^l v$ (Logemann and Ryan 2014, Theorem 2.11) then yields the desired result. In this setting it is even possible to explicitly describe what “close” to the equilibrium means:

Denote by $m(\lambda_j)$ the multiplicity of λ_j and by $(v_j^0, \dots, v_j^{m(\lambda_j)-1})$ the basis of the generalized eigenspace $\text{Eig}(\lambda_j) := \ker((Q^d)^T - \lambda_j I)^{m(\lambda_j)}$. Choose $\epsilon > 0$ such that for all $\bar{m} \in N_\epsilon(\bar{m})$ it holds that $\mathcal{D}(\bar{m}) = \{d\}$ and define

$$C_j^k := \sum_{l=0}^{m(\lambda_j)-1} e^{-l} \frac{l^l}{l!(-\operatorname{Re}(\lambda_j))^l} \left\| ((Q^d)^T - \lambda_j I)^l v_j^l \right\| \quad (4)$$

for all $j \in \{2, \dots, n\}$ and $k \in \{0, \dots, m(\lambda_j) - 1\}$. Then

$$\delta := \frac{\epsilon/2 \cdot \min_{j \in \{2, \dots, n\}, k \in \{0, \dots, m(\lambda_j)-1\}} \|v_i^k\|}{\min_{j \in \{2, \dots, n\}, k \in \{0, \dots, m(\lambda_j)-1\}} C_i^k}$$

is a suitable choice.

Also in the case of general dynamics we can provide a similar positive statement. However, now the evolution of the social state under the strategy d is described by the nonlinear ordinary differential equation $\dot{m} = (Q^d(m))^T m$. For this general setting we now adjust the classical condition for local stability of a nonlinear ordinary differential equation, which usually reads that all eigenvalues have negative real part, to our setting. Since $Q^d(m)$ is a generator, it is immediate that in our setting 0 is an eigenvalue of $\frac{\partial}{\partial m}(Q^d(\bar{m}))^T \bar{m}$. However, since $\dot{m} = (Q^d(m))^T m$ evolves in the probability simplex $\mathcal{P}(S)$, we obtain that the part of the solution associated to this eigenvalue and its associated eigenvector is irrelevant. Using this insight, we can prove that if the zero eigenvalue has multiplicity one and all other eigenvalues have strictly negative real parts the myopic adjustment process converges locally.

Theorem 4.3 *Let (\bar{m}, d) be a stationary mean field equilibrium such that $\mathcal{D}(\bar{m}) = \{d\}$ (that is, d is the unique optimal strategy at \bar{m}). Let $O \supseteq \mathcal{P}(S)$ be an open set such that $Q^d : O \rightarrow \mathbb{R}^{S \times S}$ is componentwise Lipschitz continuous, the matrix $Q^d(m)$ is a transition rate matrix for all $m \in \mathcal{P}(S)$ and the function $f^d : O \rightarrow \mathbb{R}^S, m \mapsto (Q^d(m))^T m$ is continuously differentiable in m . Assume further that the Jacobian $\frac{\partial}{\partial m} f^d(\bar{m})$ has a zero eigenvalue with eigenvector \bar{m} and all other eigenvalues have strictly negative real parts. Then there is a $\delta > 0$ such that any solution of the myopic adjustment process (3) with initial condition $m_0 \in \mathcal{P}(S) \cap N_\delta(\bar{m})$ converges exponentially fast to \bar{m} , i.e. there are constants $C_1, C_2 > 0$ such that*

$$\|m(t) - \bar{m}\| \leq C_1 e^{-C_2 t} \quad \text{for all } t \geq 0.$$

Remark 4.4 This result covers Theorem 4.2, since the Jacobian matrix in the case of constant dynamics is $(Q^d)^T$. However, the proof is non-constructive. In particular, in contrast to the setting of Theorem 4.2, we cannot explicitly describe δ .

These theorems can be directly applied in examples: Indeed, we obtain for the consumer choice model introduced in Sect. 3 and analysed in Neumann (2020) that local convergence happens to any deterministic stationary equilibrium where the equilibrium distribution does not equal the boundary value k_1 or k_2 , respectively. Also in corruption and botnet defence models (see Kolokoltsov and Malafeyev 2017; Kolokoltsov and Bensoussan 2016) usually most stationary equilibria have a deterministic equilibrium strategy that is unique for the equilibrium point. Hence, they form a natural class to apply these results. For example in a simplified version

of the corruption model presented in Kolokoltsov and Malafeyev (2017), we can show, relying on the analysis in Neumann (2020), that for any parameter choice there is local convergence towards those deterministic stationary equilibria where the equilibrium distribution lies in the interior of $\mathcal{P}(\mathcal{S})$. Moreover, for some parameter constellations we also obtain local convergence towards the deterministic stationary equilibria where the equilibrium distribution is such that all agents are corrupt or all agents are honest.

Let us finally comment on the local stability of the myopic adjustment process for equilibria (\bar{m}, π) such that $|\mathcal{D}(\bar{m})| > 1$, i.e. in the case that there are more than one (indeed infinitely many) optimal strategies for \bar{m} . In this case a necessary condition for local stability would be that under any strategy $\pi \in \text{conv}(\mathcal{D}(\bar{m}))$ the solution of the ordinary differential equation $\dot{m} = (Q^\pi(m))^T m$ converges to \bar{m} . However, this would in particular mean, that $0 = Q^\pi(\bar{m})^T \bar{m}$ for all $\pi \in \text{conv}(\mathcal{D}(\bar{m}))$. This means that local stability results would be only possible for the case that (\bar{m}, π) is an equilibrium for all $\pi \in \text{conv}(\mathcal{D}(\bar{m}))$, a special case that will not often occur. Since the results and conditions would be in the same spirit as Theorems 4.2 and 4.3, we do not present any local stability results for non-strict equilibria here.

5 Global convergence

The question of global convergence “Given an arbitrary initial condition $m_0 \in \mathcal{P}(\mathcal{S})$ does any trajectory converge towards some mean field equilibrium?” is much more complex. Here, we prove two global convergence results, which do not directly yield the desired convergence statement, instead we only obtain convergence towards equilibria with a deterministic equilibrium strategy or that the trajectory remains in a set where at least two deterministic strategies are simultaneously optimal. However, relying on example-specific properties, we can then often prove the convergence towards the mixed strategy equilibria by hand. This idea is illustrated in Example 5.2.

In order to prove these statements we again have to assume that the myopic adjustment process can be defined on an open superset \mathcal{O} of $\mathcal{P}(\mathcal{S})$: For this we assume that the functions Q_{ija} and r_{ia} are still Lipschitz continuous on this set for all $i, j \in \mathcal{S}$ and $a \in \mathcal{A}$ and that $Q_{\cdot a}(m)$ is a generator matrix for all $a \in \mathcal{A}$ and $m \in \mathcal{O}$. Then we write $\mathcal{D}(m)$ for the set of all optimal deterministic stationary strategies for the continuous time Markov decision process with transition rates $Q_{ija}(m)$ and reward function $r_{ia}(m)$ and define $F : \mathcal{O} \rightarrow 2^{\mathbb{R}^S}$ by

$$F(m) := \text{conv} \left\{ \left(\sum_{i \in \mathcal{S}} \sum_{a \in \mathcal{A}} m_i Q_{ija}(m) d_{ia} \right) : d \in \mathcal{D}(m) \right\}.$$

Let us moreover introduce the set

$$\mathcal{U} := \{d \in D^s : \mathcal{D}(m) = \{d\} \text{ for some } m \in \mathcal{P}(\mathcal{S})\}$$

that collects all those strategies that are for some social state the unique optimal strategy. Note that a strategy that satisfies $d \in \mathcal{D}(m)$ for some $m \in \mathcal{P}(\mathcal{S})$ does not automatically lie in \mathcal{U} . It only does this, if there is an $m \in \mathcal{P}(\mathcal{S})$ such that $\mathcal{D}(m) = \{d\}$. Therefore, the restriction that $|\mathcal{U}| = 2$, which we will consider in the first part of this section, is not as strict as it seems, see Example 5.2.

As a final preparation we review nonlinear Markov chains, which naturally arise in our context and play a crucial role for the assumptions of our theorems: Namely, whenever $\mathcal{D}(m) = \{d\}$ the trajectory of the myopic adjustment process equals the evolution of the marginals of a nonlinear Markov chain with generator $Q^d(\cdot)$. Thus, the long-term behaviour of the myopic adjustment process is partly governed by the long-term behaviour of the nonlinear Markov chains with generator $Q^d(\cdot)$, $d \in \mathcal{D}^s$. Nonlinear Markov chains are a generalization of classical Markov chains with the new feature that the transition probabilities do not only depend on the current state, but also on the current distribution of the process. Thus, the processes are characterized through the transition probabilities $(P_{ij}(t, m))_{i,j \in \mathcal{S}}$, which describe the probability to be in state j and time t when at time 0 the state was i and the initial distribution was m , or (non-uniquely) through the marginal distributions $\Phi_i^t(m)$, which describes the probability to be in state i at time t when the initial distribution was m . One can show that it is indeed sufficient to characterize a nonlinear Markov chain through a nonlinear generator, that is a Lipschitz continuous function $Q : \mathcal{P}(\mathcal{S}) \rightarrow \mathbb{R}^{\mathcal{S} \times \mathcal{S}}$ such that $Q(m)$ is a conservative generator for all $m \in \mathcal{P}(\mathcal{S})$. As in the theory of standard Markov chains, the invariant distribution is a central tool for the analysis of the long-term behaviour and it solves the (now non-linear) equation $0 = Q(m)^T m$. For our purpose it suffices to require that the nonlinear Markov chain converges in the limit towards some invariant distribution, that is that for all $m_0 \in \mathcal{P}(\mathcal{S})$ there is an invariant distribution $\bar{m}(m_0)$ such that

$$\lim_{t \rightarrow \infty} \|\Phi^t(m_0) - \bar{m}(m_0)\| = 0.$$

We remark that this condition is weaker than ergodicity (see Neumann 2023, Section 4.2). (For more details consider Kolokoltsov (2010) and (in particular regarding the long-term behaviour) Neumann (2023).)

We will now first provide a simple global convergence statement for the special case where $|\mathcal{U}| = 2$, in which case the ideas used in the general proof become clearer. Thereafter, we provide a general statement.

5.1 Global convergence for $|\mathcal{U}| = 2$

If $\mathcal{U} = \{d^1, d^2\}$ then the differential inclusion (3) describing the myopic adjustment process simplifies substantially: Define

$$\begin{aligned} g(m) &:= \left(V^{d^2}(m) - V^{d^1}(m) \right) \cdot 1 \\ &= \left((\beta I - Q^{d^2}(m))^{-1} r^{d^2}(m) - (\beta I - Q^{d^1}(m))^{-1} r^{d^1}(m) \right) \cdot 1, \end{aligned}$$

where $V^{d^1}(m)$ and $V^{d^2}(m)$ are the reward vectors of the Markov decision process with transition rates $Q_{ija}(m)$ and $r_{ia}(m)$. Moreover, assume that g is twice continuously differentiable and Lipschitz continuous. Since $V^{d^1}(m) \geq V^{d^2}(m)$ or $V^{d^1}(m) \leq V^{d^2}(m)$ for all $m \in \mathcal{P}(\mathcal{S})$ by classical results on Markov decision processes it is immediate that $g(m) < 0$ if and only if $\mathcal{D}(m) = \{d^1\}$, $g(m) = 0$ if and only if $\mathcal{D}(m) = \{d^1, d^2\}$ and $g(m) > 0$ if and only if $\mathcal{D}(m) = \{d^2\}$. This means that $g(m)$ describes the behaviour of agents given the social state $m \in \mathcal{O}$. Therefore, we have

$$F(m) := \begin{cases} \left(\sum_{i \in \mathcal{S}} m_i Q_{ij}^{d^1}(m) \right)_{j \in \mathcal{S}} & g(m) < 0 \\ \text{conv} \left\{ \left(\sum_{i \in \mathcal{S}} m_i Q_{ij}^{d^1}(m) \right)_{j \in \mathcal{S}}, \left(\sum_{i \in \mathcal{S}} m_i Q_{ij}^{d^2}(m) \right)_{j \in \mathcal{S}} \right\} & g(m) = 0 \\ \left(\sum_{i \in \mathcal{S}} m_i Q_{ij}^{d^2}(m) \right)_{j \in \mathcal{S}} & g(m) > 0 \end{cases}.$$

Theorem 5.1 Assume that for all $m \in \mathcal{O}$ such that $g(m) = 0$ it holds that $\nabla g(m) \neq 0$. Furthermore, assume that the nonlinear Markov chains with transition rate matrix functions $Q^{d^1}(m)$ and $Q^{d^2}(m)$ converge in the limit towards some stationary distribution.

- (i) Assume that for all $m \in \mathcal{O}$ such that $g(m) = 0$ it holds that

$$\langle (Q^{d^1}(m))^T m, \nabla g(m) \rangle > 0 \quad \text{and} \quad \langle (Q^{d^2}(m))^T m, \nabla(-g)(m) \rangle < 0.$$

Then the myopic adjustment process converges towards some stationary mean field equilibrium with deterministic equilibrium strategy from \mathcal{U} .

- (ii) Assume that for all $m \in \mathcal{O}$ such that $g(m) = 0$ it holds that

$$\langle (Q^{d^1}(m))^T m, \nabla g(m) \rangle < 0 \quad \text{and} \quad \langle (Q^{d^2}(m))^T m, \nabla(-g)(m) \rangle > 0.$$

Then the myopic adjustment process converges towards some stationary mean field equilibrium with deterministic equilibrium strategy from \mathcal{U} .

- (iii) Assume that for all $m \in \mathcal{O}$ such that $g(m) = 0$ it holds that

$$\langle (Q^{d^1}(m))^T m, \nabla g(m) \rangle \geq 0 \quad \text{and} \quad \langle (Q^{d^2}(m))^T m, \nabla(-g)(m) \rangle \geq 0.$$

Then the myopic adjustment process either converges towards a deterministic stationary mean field equilibrium with equilibrium strategy from \mathcal{U} or there is a $T > 0$ such that the process satisfies $g(m(t)) = 0$ for all $t > T$.

The gradient conditions in the theorem now link for the case that both strategies are simultaneously optimal (i.e. $g(m) = 0$) the evolution of the strategic behaviour of the individual agents ($\nabla g(m)$) with the evolution of the social state given the deterministic strategies d^1 and d^2 (given by $(Q^{d^1}(m))^T m$ and $(Q^{d^2}(m))^T m$, respectively). The conditions state that this evolution should be in the same

“direction” for all points $m \in \mathcal{P}(\mathcal{S})$ that satisfy $g(m) = 0$. Moreover, they explicitly encode the direction to which the social state moves under the myopic adjustment process: The conditions in case (i) state that when $g(m) = 0$ the social state heads to the set where the strategy d^2 is optimal, and the conditions in case (ii) state that when $g(m) = 0$ the social state heads to the set where the strategy d^1 is optimal. In case (iii) the conditions say, that when $g(m) = 0$ and the population chooses d^1 the distribution tends into the set where d^2 is optimal, and, when $g(m) = 0$ and the population chooses d^2 the distribution tends into the set where d^1 is optimal. Therefore, the trajectory will always stay in the set where both strategies are simultaneously optimal.

The proof of the theorem relies on the fact that the shape of the myopic adjustment process F does only depend on the fact whether $g(m) = 0$, $g(m) < 0$ or $g(m) > 0$. Depending on the gradient conditions we prove for each of the three cases that once $g(m(t)) > 0$ or $g(m(t)) < 0$ or $g(m(t)) = 0$, respectively, the function $g(m(\cdot))$ will be greater or less or equal zero for all future times. This yields, together with the fact that $Q^{d^1}(\cdot)$ and $Q^{d^2}(\cdot)$ are converging in the limit, to the desired convergence result.

Example 5.2 Let us consider the following example, which consists of two “good” states, where a positive reward is earned, and one “bad” state, where no reward is earned. The agents in the “good” state face congestion effects, namely there is a risk, increasing in the share of individuals in that state, to go to the “bad” state. The control options are to switch between the two good states. One can interpret this model as a stylized model of the choice between two mobile phone providers, where the customer faces the risk of a breakdown in connection that increases in the share of customers using the same provider. For simplicity, we assume that agents in the “bad” state have no choice option, but recover into each of the two states with equal probability.

The formal characterization is given by $\mathcal{S} = \{1, 2, 3\}$ and $\mathcal{A} = \{change, stay\}$ together with

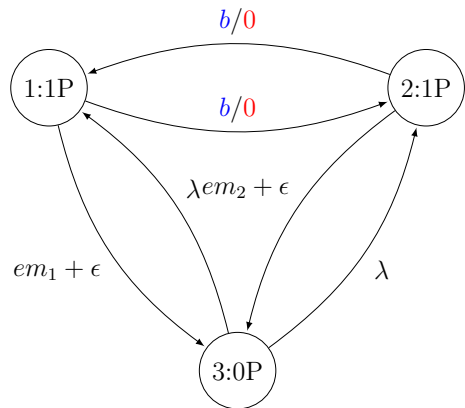
$$Q_{\cdot change}(m) = \begin{pmatrix} -(b + em_1 + \epsilon) & b & em_1 + \epsilon \\ b & -(b + em_2 + \epsilon) & em_2 + \epsilon \\ \lambda & \lambda & -2\lambda \end{pmatrix}$$

$$Q_{\cdot stay}(m) = \begin{pmatrix} -(em_1 + \epsilon) & 0 & em_1 + \epsilon \\ 0 & -(em_2 + \epsilon) & em_2 + \epsilon \\ \lambda & \lambda & -2\lambda \end{pmatrix}$$

and $r_{\cdot stay}(m) = r_{\cdot change}(m) = (1, 1, 0)$, where all constants are strictly positive. A visualization of the model is given in Fig. 2.

In Neumann (2019) it is shown that there are infinitely many mixed strategy equilibria with equilibrium distribution

Fig. 2 Representation of Example 5.2



$$\left(\frac{\sqrt{4\lambda^2 + 8\lambda\epsilon + \epsilon^2} - 2\lambda - \epsilon}{2\epsilon}, \frac{\sqrt{4\lambda^2 + 8\lambda\epsilon + \epsilon^2} - 2\lambda - \epsilon}{2\epsilon}, \frac{4\epsilon + 4\lambda - \sqrt{4\lambda^2 + 8\lambda\epsilon + \epsilon^2}}{2\epsilon} \right)$$

with equilibrium strategies satisfying $\pi_{1,change} = \pi_{2,change}$.

To apply Theorem 5.1 we first note, that in Neumann (2023) it is shown that the relevant Markov chains are strongly ergodic and thus, converge to some limit distribution. Moreover, choosing $\mathcal{U} = \{(change, stay), (stay, change)\}$, $O = (-\frac{\epsilon}{b}, \infty) \times (-\frac{\epsilon}{b}, \infty) \times \mathbb{R}$ and $g(m) = m_1 - m_2$ we obtain that

$$\begin{aligned} \langle Q^{cs}(m)^T m, \nabla - g(m) \rangle &= 2bm_2 \geq 0 \\ \langle Q^{sc}(m)^T m, \nabla g(m) \rangle &= 2bm_2 \geq 0. \end{aligned}$$

Thus, Theorem 5.1 yields that either convergence towards a stationary equilibrium with an equilibrium strategy from \mathcal{U} happens or that there is a $T \geq 0$ such that $g(m(t)) = 0$ for all $t \geq T$. Since there is no stationary equilibrium with an equilibrium strategy from \mathcal{U} it is clear that there is a $T \geq 0$ such that $g(m(t)) = 0$ for all $t \geq T$, which means that $m_1(t) = m_2(t)$ for all $t \geq T$. Thus, also $\dot{m}_1(t) = \dot{m}_2(t)$. By (3), this yields that

$$\begin{aligned} & -\pi_{1,change}(t)bm_1(t) - \epsilon m_1(t)^2 - \epsilon m_1(t) + \pi_{2,change}(t)bm_1(t) + \lambda m_3(t) \\ & = \pi_{1,change}(t)bm_1(t) - \pi_{2,change}(t)bm_1(t) - \epsilon m_1(t)^2 - \epsilon m_1(t) + \lambda m_3(t), \\ \text{i.e. } & -\pi_{1,change}(t)bm_1(t) + \pi_{2,change}(t)bm_1(t) = 0. \end{aligned}$$

Thus, for almost all $t \geq T$ the trajectory of the myopic adjustment process has to satisfy

$$\dot{m}_1(t) = -\epsilon m_1(t)^2 - (\epsilon + 2\lambda)m_1(t) + \lambda,$$

which is a Riccati equation, for which $[0, 1]$ is flow invariant and for which a unique classical solution for any initial condition $m_0 \in [0, 1]$ exists. Numerical simulations

indicate that in our setting with initial conditions $m_0 \in [0, 1]$ convergence towards the distribution of the stationary mixed strategy equilibria is likely. \square

5.2 Global convergence for the general case

The idea in the simple two strategy case was that the set of social states, where d^1 or d^2 or both strategies, respectively, are optimal is characterized by the function g . Describing the behaviour in a neighbourhood of $\{m \in O : g(m) = 0\}$ by gradient conditions allows understand that once $g(m(t)) = 0$ it will not happen again (in case (i) or (ii)) or $g(m(t))$ will remain at 0 (in the case (iii)) for all future times, which then allows to understand the exact behaviour of the myopic adjustment process from this time on. Here, we want use a similar approach: First, we define a function g_d such that $g_d(m) < 0$ means that d is the unique optimal strategy for m ($\mathcal{D}(m) = d$), $g_d(m) = 0$ means that d is one of multiple optimal strategies ($|\mathcal{D}(m)| \geq 2, d \in \mathcal{D}(m)$) and $g_d(m) > 0$ means that d is not optimal for m . Thereafter, we set up gradient conditions for all those $m \in O$ satisfying $g_d(m) = 0$ that again describe consistent behaviour.

Assumption A2 For each $m \in O$ there is a strategy $d \in \mathcal{U}$ such that $d \in \mathcal{D}(m)$.

Moreover, let us write $\text{Opt}^{\text{unique}}(d)$ for the set of all $m \in O$ such that $\mathcal{D}(m) = \{d\}$ and $\text{Opt}^{\text{some}}(d)$ for the set of all $m \in O$ such that $d \in \mathcal{D}(m)$. For these sets let us assume the following:

Assumption A3 For each strategy $d \in \mathcal{U}$ there exists a twice continuously differentiable, Lipschitz continuous function $g_d : O \rightarrow \mathbb{R}$ such that

$$\begin{aligned}\text{Opt}^{\text{unique}}(d) &= \{m \in O : g_d(m) < 0\} \\ \text{Opt}^{\text{some}}(d) &= \{m \in O : g_d(m) \leq 0\}\end{aligned}$$

and such that for all $d \in \mathcal{U}$ and all $m \in \{m \in O : g_d(m) = 0\}$ we have $\nabla g_d(m) \neq 0$.

Assumption A4 For any $m \in \mathcal{P}(\mathcal{S})$ there are at most two strategies $d \in \mathcal{U}$ such that $m \in \text{Opt}^{\text{some}}(d)$.

Assumption A5 The set $\mathcal{P}(\mathcal{S})$ is flow invariant for $\dot{m} \in F(m)$.

Assumption A6 For all $d \in \mathcal{U}$ and all $d^1, d^2 \in \mathcal{U}$ satisfying $\text{Opt}^{\text{some}}(d^1) \cap \text{Opt}^{\text{some}}(d^2) \neq \emptyset$ we have

$$\langle (Q^d(m))^T m, \nabla g_{d^1}(m) \rangle \geq 0 \quad \text{for all } m \in \text{Opt}^{\text{some}}(d^1) \cap \text{Opt}^{\text{some}}(d^2)$$

or

$$\langle (Q^d(m))^T m, \nabla g_{d^1}(m) \rangle \leq 0 \quad \text{for all } m \in \text{Opt}^{\text{some}}(d^1) \cap \text{Opt}^{\text{some}}(d^2).$$

Let us briefly comment on the assumptions: The Assumptions A2, A3 and A4 are automatically satisfied in the previously described case with $|\mathcal{U}| = 2$. The fifth assumption could be replaced by Assumption A6 with strict inequalities for $m \in \mathcal{P}(\mathcal{S})$ with $m_i = 0$ for some $i \in \mathcal{S}$ (see Appendix B). However, this would make the statement less general, since for example Assumption A5 is satisfied in Example 5.2, but Assumption A6 with strict inequalities for $m \in \mathcal{P}(\mathcal{S})$ with $m_i = 0$ for some $i \in \mathcal{P}(\mathcal{S})$ is not satisfied.

The Assumption A6 captures partly the conditions imposed in the three different cases of Theorem 5.1. The rest of these conditions will be described by the family of digraphs that helps to characterize the long-term behaviour of the myopic adjustment process: For any $d \in \mathcal{U}$ let $D(d)$ be a digraph with vertex set \mathcal{U} and arc set given by $d^1 \rightarrow d^2 \in A(D(d))$ if and only if $\text{Opt}^{\text{some}}(d^1) \cap \text{Opt}^{\text{some}}(d^2) \neq \emptyset$ and

$$\langle (Q^d(m))^T m, \nabla g_{d^1}(m) \rangle > 0 \quad \text{for all } m \in \text{Opt}^{\text{some}}(d^1) \cap \text{Opt}^{\text{some}}(d^2).$$

This digraph can be interpreted as follows: If $d^1 \rightarrow d^2$ then the nonlinear Markov chain with generator $Q^d(\cdot)$ will when it hits the set $\text{Opt}^{\text{some}}(d^1) \cap \text{Opt}^{\text{some}}(d^2)$ tends to the set $\text{Opt}^{\text{some}}(d^2)$.

If the digraph $D(d)$ is acyclic then it provides a good description of the behaviour of solutions of $\dot{m} = (Q^d(m))^T m$. Namely, by a well known result from graph theory there is in this case an acyclic ordering, which is an ordering $d^1 \leq \dots \leq d^u$ of the vertices of $D(d)$ such that whenever $d^i \leq d^j$ there is no edge from d^j to d^i . For this acyclic ordering one can prove that for any strategy $k \in \{1, \dots, u\}$ the set $\mathcal{P}(\mathcal{S}) \cap \bigcup_{l \geq k} \text{Opt}^{\text{some}}(d^l)$ is flow invariant, i.e. once the trajectory hits this set, it will never leave it. Loosely speaking, this means that the trajectory moves through the optimality sets $\text{Opt}^{\text{some}}(d^l)$ in a way that is consistent with the arrows in the digraph $D(d)$.

Before we present the main theorem of the section, we verify that the consumer choice model introduced in Example 3.4 satisfies all assumptions:

Example 5.3 We choose $O = (-\delta, 1 + \delta)^2$, $\mathcal{U} = \{cs, ss, sc\}$ and setting

$$\begin{aligned} g_{cs}(m) &= m_1 - k_1 \\ g_{ss}(m) &= -(m_1 - k_1) \cdot (k_2 - m_1) \\ g_{sc}(m) &= k_2 - m_1, \end{aligned}$$

completes the set-up. Moreover, noting that $k_1 < k_2$ and

$$\begin{aligned}\nabla g_{cs}(m) &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} && \Rightarrow \nabla g_{cs}(k_1, m_2) \neq 0 \\ \nabla g_{cs}(m) &= \begin{pmatrix} k_2 + k_1 - 2m_1 \\ 0 \end{pmatrix} && \Rightarrow \nabla g_{ss}(k_1, m_2) \neq 0 \\ &&& \text{and } \nabla g_{ss}(k_2, m_2) \neq 0 \\ \nabla g_{sc}(m) &= \begin{pmatrix} -1 \\ 0 \end{pmatrix} && \Rightarrow \nabla g_{sc}(k_1, m_2) \neq 0,\end{aligned}$$

it is immediate that Assumptions A2, A3 and A4 are satisfied. Assumption A5 is immediate since in a neighbourhood of $(1, 0)$ and $(0, 1)$ we face a classical ODE that can be solved explicitly and for which moreover $\mathcal{P}(\mathcal{S})$ is flow invariant. Moreover, also Assumption A6 is satisfied since the sets $\text{Opt}^{\text{some}}(d^1) \cap \text{Opt}^{\text{some}}(d^2)$ are singletons.

With all these preparations let us now formulate the main theorem of this section:

Theorem 5.4 *Let Assumptions A1, A2, A3, A4, A5 and A6 hold and assume that:*

- (i) *For all $d \in \mathcal{U}$ the nonlinear Markov chain with transition rate matrix function $Q^d(\cdot)$ converges in the limit to some stationary distribution.*
- (ii) *For all $d \in \mathcal{U}$ the digraph $D(d)$ is acyclic.*
- (iii) *There exists an ordering d^1, d^2, \dots, d^u of \mathcal{U} such that for each $i \in \{1, \dots, u\}$ there exists an acyclic ordering $\leq_{D(d^i)}$ of $D(d^i)$ such that*

$$d^i \leq_{D(d^i)} d^j \quad \text{for all } j \geq i.$$

- (iv) *If every ordering that satisfies the conditions of (iii) has the same final vertex \hat{d} , then for any $d \in \mathcal{U} \setminus \{\hat{d}\}$ either $\text{Opt}^{\text{some}}(d) \cap \text{Opt}^{\text{some}}(\hat{d}) = \emptyset$ or $\hat{d} \rightarrow d \notin D(\hat{d})$ with*

$$\langle (Q^d(m))^T m, \nabla g_d(m) \rangle < 0 \quad \text{for all } m \in \text{Opt}^{\text{some}}(d) \cap \text{Opt}^{\text{some}}(\hat{d}).$$

Then for any initial condition $m_0 \in \mathcal{P}(\mathcal{S})$ the myopic adjustment process either converges towards the distribution \bar{m} of some deterministic mean field equilibrium (\bar{m}, d) or there is a $T > 0$ and a strategy $d \in \mathcal{U}$ such that the trajectory stays in

$$\mathcal{P}(\mathcal{S}) \cap (\text{Opt}^{\text{some}}(d) \setminus \text{Opt}^{\text{unique}}(d))$$

for all $t \geq T$.

Let us first comment that all four conditions can be intuitively justified: Condition (i) guarantees that whenever we stay inside an optimality set $\text{Opt}^{\text{unique}}(d)$ for all subsequent times then we converge towards a deterministic stationary mean

field equilibrium. Condition (ii) ensures that the evolution of the social state given strategy d^i is consistent on all sets where two strategies are simultaneously optimal, i.e. the evolution tends for all points to the set where a particular strategy is optimal. Moreover, the condition ensures that the evolution is not cyclic in the sense that it moves through a sequence of optimality sets again and again. Condition (iii) then links like the gradient conditions in Theorem 5.1 the evolution of the strategic behaviour with the evolution of the social state. First, it requires that for any pair of strategies that are simultaneously optimal the behaviour is locally as specified in Theorem 5.1. Second, it also requires that the behaviour of the agents on all these sets where two strategies are simultaneously optimal is consistent, in the sense that the behaviour is not cyclic as described before. The mainly technical condition (iv) ensures that for the final vertex of the acyclic ordering we will never jump from $\text{Opt}^{\text{unique}}(d)$ to $\text{Opt}^{\text{some}}(d) \setminus \text{Opt}^{\text{unique}}(d)$ and back infinitely often, which is necessary to ensure the desired convergence.

The conditions (iii) and (iv) seem to be rather complex. However, relying on the well-known fact that a digraph is acyclic if and only if there is an acyclic ordering of its vertices as well the classical algorithm to obtain such a sequence, it is also possible to verify conditions (iii) and (iv) using a (polynomial-time) algorithm. This is in detail explained in Appendix A.

If we compare the conditions here with the conditions of Theorem 5.1, we see that condition (i) of Theorem 5.4 is also present in the conditions of Theorem 5.1. The other three conditions (ii)–(iv) are equivalent to the three distinct cases covered in Theorem 5.1. Such a case distinction is however, not sensible in the context of a larger number of strategies, for which reason we utilize the formalization by the digraphs $D(d)$.

To conclude the section, let us apply Theorem 5.4 in the consumer choice model introduced in Example 3.4.

Example 5.5 We already verified in Example 5.3 that the model satisfies Assumptions A1–A6. Moreover, condition (i) of Theorem 5.4 is satisfied since $Q^d(m)$ is a standard Markov chain with irreducible generator. Condition (ii) is satisfied since we can solve the differential equation $\dot{m} = (Q^d(m))^T m$ explicitly and obtain that these solutions are monotone in m_1 . Using the algorithm from Sect. A we obtain that conditions (iii) and (iv) are only satisfied in those three of eight cases discussed in Neumann (2020) where a unique equilibrium exists. In these cases global convergence towards this unique equilibrium is obtained since the solution cannot remain in the sets $\{(k_1, 1 - k_1)\}$ or $\{(k_2, 1 - k_2)\}$ for all $t \geq T$ for some $T \geq 0$. In the other five cases the condition (iii) is not satisfied since we obtain that in the sets $\text{Opt}^{\text{some}}(sc) \cap \text{Opt}^{\text{some}}(cc)$ or $\text{Opt}^{\text{some}}(cs) \cap \text{Opt}^{\text{some}}(cc)$ the trajectories of the myopic adjustment process can evolve non-uniquely either it moves towards $\text{Opt}^{\text{unique}}(sc)$ or it moves towards $\text{Opt}^{\text{unique}}(ss)$ or it stays in $\text{Opt}^{\text{some}}(cs) \cap \text{Opt}^{\text{some}}(cc)$ (or with cs replaced by sc). In the current two-dimensional setting with linearly ordered optimality sets we still observe convergence (which however cannot be expected in

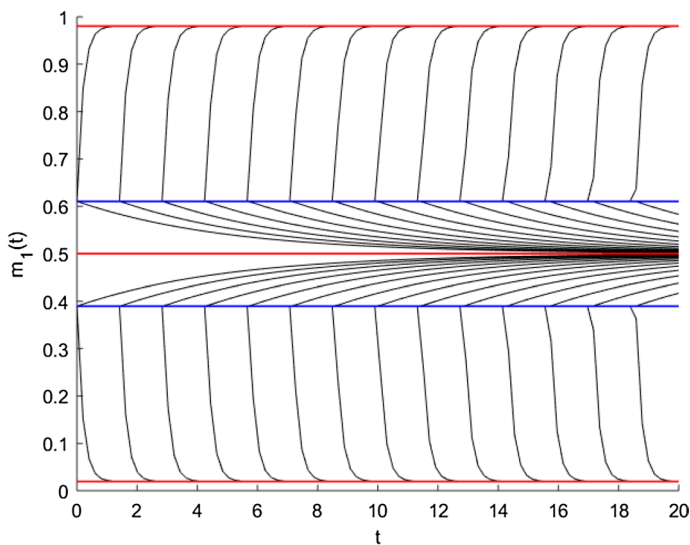


Fig. 3 Some of the infinitely many possible solutions with initial condition $(m_0)_1 = k_1$ and $(m_0)_1 = k_2$. Additionally, the red vertical lines depict the stationary points given the strategies *cs*, *ss* and *sc* (from bottom to top) and the blue vertical lines depict the crucial thresholds k_1 and k_2

general). However, for the starting point $(k_1, 1 - k_1)$ or $(k_2, 1 - k_2)$ the long-term behaviour is somewhat unstable. Indeed, the trajectory can remain at $(k_1, 1 - k_1)$ (or $(k_2, 1 - k_2)$ respectively) until time $T \in [0, \infty]$ and thereafter the process can either converge to the equilibrium given the strategy *sc* (or *cs* respectively) or it converges to the equilibrium given the strategy *ss*. This behaviour is illustrated in Fig. 3.

6 Conclusion

This paper introduces a learning procedure for mean field games with finite state and action space. More precisely, at any time the agents assume that the social state is constant and choose the optimal strategy given this social state. The learning procedure is non-standard since it involves two layers of adjustment - the adjustment of the strategies and the evolution of the individual states. Yet, we obtain local and global stability results that are similar to classical results from evolutionary game theory: We show that strict equilibria are locally stable if the dynamics have constant transition rates that form an irreducible generator or if the dynamics are a nonlinear sink on the probability simplex. Moreover, we prove global convergence under assumptions that ensure that at each point at most two strategies primarily influence the evolution of the myopic adjustment process and that at each point the local payoff structure and transition rates guide the agent consistently to some direction.

Appendix A. An algorithm to verify the consistency condition

The question whether an ordering satisfying condition (iii) in Theorem 5.4 exists or not seems to be complex at first sight. One would have to check for all possible permutations whether the ordering satisfies the condition. However, the close connection to the notion of acyclic orderings allows to provide a polynomial algorithm that determines whether such an ordering exists or not. As in the case of acyclic orderings, it is moreover possible to formulate a polynomial time algorithm to find all orderings satisfying (iii). Due to the additional notational complexity, we omit this here.

Our algorithm is a modification of the following simple algorithm to determine an acyclic ordering if it exists: Namely, in each step a vertex with indegree 0 is picked, added to the tail of the acyclic ordering obtained so far and deleted from the digraph (Bang-Jensen and Gutin 2010, Section 2.1). Since we want to construct an ordering $\{d^1, \dots, d^S\}$ such that $\{d^{i+1}, \dots, d^S\}$ lies behind d^i in an acyclic ordering of $D(d^i)$ for each strategy $d^i \in \mathcal{U}$ the central modification of the algorithm is that when we add a vertex to the ordering we do not only delete the vertex and its arcs from the digraphs, but that we also add arcs in order to ensure that an acyclic ordering of $D(d) \setminus \{d^1, \dots, d^i\}$ is also an acyclic ordering of $D(d)$. More precisely, if we add d to the ordering since it had indegree 0 in $D(d)$, then we add the arcs $d^1 \rightarrow d^2$ to the graphs $D(\hat{d})$ for all $\hat{d} \in \hat{V}$ whenever $d^1 \rightarrow d$ and $d \rightarrow d^2$ are both arcs in $D(\hat{d})$. This yields that two vertices $d^1, d^2 \in \hat{V}$ are connected in the modified digraph if and only if they are connected in the original graph $D(\hat{d})$. This algorithm is formalized in Algorithm 1 and Algorithm 2.

Algorithm 1: GraphModification

Data: A digraph $D = (V, A)$ and a vertex x

Result: A new digraph $\hat{D} = (\hat{V}, \hat{A})$ such that $\hat{V} = V \setminus \{x\}$ and for all $x^1, x^2 \in \hat{V}$ there is a path from x^1 to x^2 in D if and only if there is a path from x^1 to x^2 in \hat{D}

```

1  $\hat{V} \leftarrow V \setminus \{x\}$ 
2  $\hat{A} \leftarrow A$ 
3 for  $x^1$  such that  $x^1 \rightarrow x \in A$  do
4   for  $x^2$  such that  $x \rightarrow x^2 \in A$  do
5      $\quad$  add  $x^1 \rightarrow x^2$  to  $\hat{A}$ 
6 delete all arcs containing  $x$  from  $\hat{A}$ 
```

Algorithm 2: An Algorithm to Find an Acyclic Ordering (if it exists)

Data: A family of digraphs $(D(d))_{d \in \mathcal{U}}$
Result: An ordering (d^1, \dots, d^u) of \mathcal{U} satisfying (iii) or \emptyset (if no such ordering exists)

```

1  $\hat{V} \leftarrow \mathcal{U}$ 
2  $I \leftarrow$  set of all  $d \in \mathcal{U}$  with indegree 0 in  $D(d)$ 
3  $O \leftarrow$  an empty list
4 while  $I$  non-empty do
5   remove an element  $d$  from  $I$ 
6   remove  $d$  from  $\hat{V}$ 
7   add  $d$  to the tail of  $O$ 
8   for  $\hat{d} \in \hat{V}$  do
9      $D(\hat{d}) \leftarrow \text{GraphModification}(D(\hat{d}), d)$ 
10    if  $\hat{d}$  has no incoming arcs in  $D(\hat{d})$  then
11      add  $\hat{d}$  to  $I$ 
12 if  $\hat{V}$  is non-empty then
13   return  $\emptyset$ 
14 else
15   return  $O$ 

```

The proof that the algorithm works as desired relies on the following two easy-to-verify properties of Algorithm 1:

Lemma A.1 *Let D be a digraph, x be a vertex of D and $\hat{D} = (\hat{V}, \hat{A})$ be the digraph resulting from Algorithm 1 applied for D and x . Then the following statements hold:*

- (i) *Let $x^1, x^2 \in \hat{V}$. Then there is a path from x^1 to x^2 in D if and only if there is a path from x^1 to x^2 in \hat{D} .*
- (ii) *Let $\leq_{\hat{D}}$ be an acyclic ordering of \hat{D} . Then there is an acyclic ordering \leq of D such that $x^1 \leq x^2 \Leftrightarrow x^1 \leq_{\hat{D}} x^2$ for all $x^1, x^2 \in \hat{V}$.*
- (iii) *Let \leq_D be an acyclic ordering of D . Then there is an acyclic ordering \leq of \hat{D} such that $x^1 \leq x^2 \Leftrightarrow x^1 \leq_D x^2$ for all $x^1, x^2 \in \hat{V}$.*

Theorem A.2 *The algorithm is correct, that is whenever an ordering exists the algorithm finds one and if no ordering exists the algorithm returns \emptyset .*

Proof If the algorithm returns an ordering $O = (d^1, \dots, d^u)$, then in each step we find a strategy d^i such that d^i has indegree 0 in the digraph

$$\hat{D}(d^i) = \text{GraphModification}(\dots \text{GraphModification}(D(d^i), d^1), \dots, d^{i-1}).$$

Since d^i has indegree 0, we obtain (for example by using the standard algorithm to find an acyclic ordering, which was informally described in the beginning of this section) an acyclic ordering of $\hat{D}(d^i)$ such that $d^i \leq d^j$ for all $j > i$. By Corollary A.1,

we then also obtain an acyclic ordering of $D(d^i)$ such that $d^i \leq d^j$ for all $j \geq i$. This proves the first claim.

Now assume that (d^1, \dots, d^u) satisfies (iii). Then for each $i \in \{1, \dots, u\}$ we find an acyclic ordering of $D(d^i)$ such that $d^i \leq d^j$ for all $j > i$. In particular, Lemma A.1 yields that there is an acyclic ordering $\leq_{\hat{D}(d^i)}$ of

$$\hat{D}(d^i) = \text{GraphModification}(\dots \text{GraphModification}(D(d^i), d^1), \dots, d^{i-1})$$

such that $d^i \leq_{\hat{D}(d^i)} d^j$ for all $j > i$. However, this yields, by definition, that there is no incoming arc into d^i in $\hat{D}(d^i)$, which in particular yields that in step i the vertex d^i lies in I . Thus, choosing d^i in step i for all $1 \leq i \leq n$ yields that the algorithm outputs (d^1, \dots, d^u) . \square

Appendix B. Proofs

Proof of Lemma 3.1 Using the Kolmogorov forward equation, the individual agent's dynamics given any strategy $\pi \in \Pi^s$ can be equivalently described as being the solution of the ordinary differential equation (in the sense of Caratheodory)

$$\dot{x}_j(t) = \sum_{i \in \mathcal{S}} x_i(t) \sum_{a \in \mathcal{A}} Q_{ija}(m(t)) \pi_{ia}(t) \quad \text{for all } j \in \mathcal{S}$$

with initial condition $x(0) = x_0$. Since the aggregated strategy π of the population satisfies $\pi \in \text{conv}(\mathcal{D}(m))$ the desired claim follows. \square

Proof of Theorem 3.2 We show that the conditions of Lemma 5.1 in Deimling (1992) are satisfied, as this yields the desired existence statement. More precisely, we show in the following that

- (i) F is upper semicontinuous,
- (ii) $F(m)$ is a closed, convex set for all $m \in \mathcal{P}(\mathcal{S})$,
- (iii) there is a constant $c > 0$ such that $\|F(m)\| := \sup\{\|y\| : y \in F(m)\} \leq c(1 + \|m\|)$ for all $m \in \mathcal{P}(\mathcal{S})$, and
- (iv) $F(m) \cap T_{\mathcal{P}(\mathcal{S})}(m) \neq \emptyset$ for all $m \in \mathcal{P}(\mathcal{S})$ where

$$\begin{aligned} T_{\mathcal{P}(\mathcal{S})}(m) &= \left\{ y \in \mathbb{R}^{\mathcal{S}} : \liminf_{h \downarrow 0} \frac{d(m + hy, \mathcal{P}(\mathcal{S}))}{h} = 0 \right\} \\ &= \left\{ y \in \mathbb{R}^{\mathcal{S}} : y_i \geq 0 \forall i \in \mathcal{S} \text{ s.t. } m_i = 0 \wedge \sum_{i \in \mathcal{S}} y_i = 0 \right\} \end{aligned}$$

(Aubin and Cellina 1984, Proposition 5.1.7).

(i): Let $m \in \mathcal{P}(\mathcal{S})$ and let $N \subseteq \mathbb{R}^{\mathcal{S}}$ be an open set such that $F(m) \subseteq N$. Since $2^{\mathcal{S}}$ equipped with the Hausdorff distance $H(\cdot, \cdot)$ is a metric space, it suffices to consider sets of the form $N = N_\epsilon(F(m))$ with $\epsilon > 0$. Since $\mathcal{D}(m) = \{d \in D^s : V^d(m) = V^*(m)\}$, we find for any $d' \in D^s \setminus \mathcal{D}(m)$ a constant

$c_{d'} > 0$ such that $V^{d'}(m) < V^* - c_{d'}$. Therefore by finiteness of D^s , the constant $c = \min_{d \in D^s \setminus \mathcal{D}(m)} c_{d'} > 0$ satisfies $V^{d'}(m) < V^*(m) - c$ for all $d' \in D^s \setminus \mathcal{D}(m)$. Since $V^d : \mathcal{P}(\mathcal{S}) \rightarrow \mathbb{R}^S$ is continuous for every $d \in D^s$, there is a $\delta_d > 0$ such that $\tilde{m} \in N_{\delta_d}(m) \Rightarrow \|V^d(\tilde{m}) - V^d(m)\| < \frac{c}{3}$. In particular, choosing $\delta_1 := \min_{d \in D^d} \delta_d$ we obtain for all $\tilde{m} \in N_{\delta_1}(m)$ and all $d' \in D^s \setminus \mathcal{D}(m)$ that pointwise

$$\begin{aligned} V^*(\tilde{m}) - V^{d'}(\tilde{m}) &\geq (V^*(\tilde{m}) - V^*(m)) + (V^*(m) - V^{d'}(m)) + (V^{d'}(m) - V^{d'}(\tilde{m})) \\ &> -\frac{c}{3} + c - \frac{c}{3} > 0. \end{aligned}$$

Thus, $d' \notin \mathcal{D}(\tilde{m})$, that is $\mathcal{D}(\tilde{m}) \subseteq \mathcal{D}(m)$. Furthermore, for $d \in D^s$ the map $F^d : \mathcal{P}(\mathcal{S}) \rightarrow \mathbb{R}^S, m \mapsto (\sum_{i \in \mathcal{S}} \sum_{a \in \mathcal{A}} m_i Q_{ija}(m) d_{ia})_{j \in \mathcal{S}}$ is continuous. Therefore, there is a $\delta_{2,d} > 0$ such that $\tilde{m} \in N_{\delta_{2,d}}(m) \Rightarrow \|F^d(\tilde{m}) - F^d(m)\| < \epsilon$. Set $\delta_2 = \min_{d \in D^s} \delta_{2,d}$. Then for $\delta := \min\{\delta_1, \delta_2\}$ it holds that $\tilde{m} \in N_\delta(m) \Rightarrow F(\tilde{m}) \subseteq N_\epsilon(F(m))$.

(ii) Since for all $m \in \mathcal{P}(\mathcal{S})$ the set $F(m)$ is a convex polytope, it is closed and convex.

(iii) Since $Q_{ija}(\cdot)$ is Lipschitz continuous for all $i, j \in \mathcal{S}$ and all $a \in \mathcal{A}$ it is moreover uniformly bounded in $m \in \mathcal{P}(\mathcal{S})$, $i, j \in \mathcal{S}$ and $a \in \mathcal{A}$ by some constant, which we denote by M . Thus, for all $d \in D^s \supseteq \mathcal{D}(m)$ we have

$$\|F^d(m)\| = \sum_{j \in \mathcal{S}} \left| \sum_{i \in \mathcal{S}} \sum_{a \in \mathcal{A}} m_i Q_{ija}(m) d_{ia} \right| \leq \sum_{j \in \mathcal{S}} \sum_{i \in \mathcal{S}} \sum_{a \in \mathcal{A}} m_i M d_{ia} = SM,$$

which implies, since $F(m)$ is the convex hull of $(F^d(m))_{d \in \mathcal{D}(m)}$ that $\|F(m)\|_1 \leq SM$.

(iv) The condition is trivially satisfied for all $m \in \text{int}(\mathcal{P}(\mathcal{S}))$ since then $T_{\mathcal{P}(\mathcal{S})}(m) = \mathbb{R}^S$ and $F(m) \neq \emptyset$ because $\mathcal{D}(m) \neq \emptyset$. Now, let $m \in \partial \mathcal{P}(\mathcal{S})$ be a boundary point. Then there is at least one $j \in \mathcal{S}$ such that $m_j = 0$. Since the only non-positive column entry of $Q_{\cdot ja}(m)$ is in row j , this implies that $F^d(m)_j \geq 0$. Moreover,

$$\sum_{j \in \mathcal{S}} F^d(m)_j = \sum_{i \in \mathcal{S}} \sum_{a \in \mathcal{A}} \underbrace{\left(\sum_{j \in \mathcal{S}} Q_{ija}(m) \right)}_{=0} d_{ia} = 0,$$

which yields that $F^d(m) \in T_{\mathcal{P}(\mathcal{S})}(m)$. Since $F(m)$ is a convex combination of $F^d(m)$ and $T_{\mathcal{P}(\mathcal{S})}(m)$ is convex, the desired claim follows. \square

Proof of Lemma 4.1 Let $V^d(m) = (V_{\delta_i}(d, m))_{i \in \mathcal{S}}$ denote the expected discounted reward vector if the player's initial individual state is i , the player chooses strategy d and the social state is m for all times. In Neumann (2020, Section 3) it is shown that $m \mapsto V^d(m)$ is continuous. Since $\mathcal{D}(\bar{m}) = \{d\}$, we have $V^d(\bar{m}) > V^{\hat{d}}(\bar{m})$ pointwise for all $\hat{d} \in D^s \setminus \{d\}$. By continuity there is an $\epsilon > 0$ such that $V^d(m') > V^{\hat{d}}(m')$ pointwise for all $m' \in N_\epsilon(\bar{m})$ and $\hat{d} \in D^s \setminus \{d\}$. Thus, $\mathcal{D}(m') = \{d\}$ for all $m' \in N_\epsilon(\bar{m})$. \square

Proof of Theorem 4.2 Choose $\epsilon > 0$ such that for all $\tilde{m} \in N_\epsilon(\bar{m})$ it holds that $\tilde{\mathcal{D}}(\tilde{m}) = \{d\}$, which is possible by Lemma 4.1. Let $(\lambda_1, \dots, \lambda_n)$ be the eigenvalues of Q^d with $\lambda_1 = 0$ and $\text{Re}(\lambda_i) < 0$ for all $i \neq 1$, which exists by Asmussen (2003, Corollary

4.9). Moreover, let us denote by $m(\lambda_j)$ the multiplicity of λ_j and by $(v_j^0, \dots, v_j^{m(\lambda_j)-1})$ the basis of the generalized eigenspace $\text{Eig}(\lambda_j) := \ker((Q^d)^T - \lambda_j I)^{m(\lambda_j)}$. We note that

$$(\bar{m}, v_2^0, \dots, v_2^{m(\lambda_2)-1}, \dots, v_n^0, \dots, v_n^{m(\lambda_n)-1})$$

is a basis of \mathbb{R}^S . Thus, for any initial condition $m_0 \in \mathbb{R}^S$ we find a unique set of constants (α_i^k) such that $m_0 = \alpha_1^0 \bar{m} + \sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k v_i^k$. By Logemann and Ryan (2014, Theorem 2.11) and since \bar{m} is the eigenvector for the eigenvalue 0 we obtain

$$m(t) = \alpha_1^0 \bar{m} + \sum_{i=2}^n e^{\lambda_i t} \sum_{l=0}^{m(\lambda_i)-1} \frac{t^l}{l!} ((Q^d)^T - \lambda_i I)^l v_i^l.$$

Since the continuous time Markov chain with generator Q^d is ergodic, we have that $m(t) \rightarrow \bar{m}$ for $t \rightarrow \infty$. Since moreover $\text{Re}(\lambda_i) < 0$ for all $i > 1$, it holds that $\alpha_1^0 = 1$. Using that the function $t \mapsto e^{\text{Re}(\lambda_i)t} \frac{t^l}{l!}$ has a unique global maximum in $[0, \infty)$ at $t = -\frac{l}{\text{Re}(\lambda_i)}$ and recalling the definition of C_j^k in (4), we obtain

$$\begin{aligned} \|m(t) - \bar{m}\| &= \left\| \sum_{i=2}^n e^{\lambda_i t} \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k \sum_{l=0}^{m(\lambda_i)-1} \frac{t^l}{l!} ((Q^d)^T - \lambda_i I)^l v_i^k \right\| \\ &\leq \sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} |\alpha_i^k| \left(\sum_{l=0}^{m(\lambda_i)-1} e^{\text{Re}(\lambda_i)t} \frac{t^l}{l!} \|((Q^d)^T - \lambda_i I)^l v_i^k\| \right) \\ &\leq \sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} |\alpha_i^k| C_i^k. \end{aligned} \quad (5)$$

If $m_0 \in N_\delta(\bar{m}) \cap \mathcal{P}(\mathcal{S})$, then

$$\|m_0 - \bar{m}\| = \sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} |\alpha_i^k| \|v_i^k\| < \frac{\epsilon \min_{i,k} \|v_i^k\|}{\max_{i,k} C_i^k},$$

which implies $\sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} |\alpha_i^k| C_i^k < \epsilon$. Therefore, $m(t) \in N_\epsilon(\bar{m})$ for all $t \geq 0$. The exponential convergence then follows from (5). \square

Proof of Theorem 4.3 The central idea of the proof has also been used in the analysis of nonlinear sinks in Hirsch and Smale (1974), namely to bound $\frac{\partial}{\partial t} \|x(t)\|_B$ with $x(t) = m(t) - \bar{m}$ for a suitable basis $B = (b_1, \dots, b_S)$ (and corresponding scalar product (i.e. $\langle b_i, b_j \rangle = \delta_{ij}$) and norm). More precisely, since for any basis B there is a matrix $C \in \mathbb{R}^{S \times S}$ such that $\langle x, y \rangle_B = x^T C y$ we obtain the product rule $\frac{\partial}{\partial t} \langle g(t), h(t) \rangle_B = \left\langle \frac{\partial}{\partial t} g(t), h(t) \right\rangle_B + \left\langle g(t), \frac{\partial}{\partial t} h(t) \right\rangle_B$ and using this

$$\frac{\partial}{\partial t} \|x(t)\|_B = \frac{\partial}{\partial t} \sqrt{\langle x(t), x(t) \rangle_B} = \frac{1}{\|x(t)\|_B} \langle \dot{x}(t), x(t) \rangle_B.$$

In the following proof we now identify, similar to Hirsch and Smale (1974), a basis B and a constant $C > 0$ such that

$$\frac{1}{\|x(t)\|_B} \langle \dot{x}(t), x(t) \rangle_B \leq -C \|x(t)\|_B,$$

from which we then deduce the desired result.

Give \mathbb{R}^S new coordinates via the transformation $x = m - \bar{m}$, which in particular means that we now consider $\tilde{f}^d(x) = f^d(x + \bar{m})$. Denote by A the Jacobian matrix of \tilde{f}^d at 0. Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of A . Without loss of generality $\lambda_1 = 0$. By assumption $\operatorname{Re}(\lambda_i) < 0$ for all $i \in \{2, \dots, n\}$, moreover, there are constants $b, c > 0$ such that $\operatorname{Re}(\lambda_i) < -b < -c$ for all $i \in \{2, \dots, n\}$.

By Hirsch and Smale (1974, Chapter 7) there is a basis

$$B = (b_1^0, b_2^0, \dots, b_2^{m(\lambda_2)-1}, \dots, b_n^0, \dots, b_n^{m(\lambda_n)-1})$$

with $b_1^0 = \bar{m}$ and corresponding inner product given by $\langle b_i^k, b_j^l \rangle = 1_{\{i=j, k=l\}}$ such that

- (i) for all $i \in \{1, \dots, n\}$ the family $(b_i^0, \dots, b_i^{m(\lambda_i)-1})$ is a basis of the generalized eigenspace $\operatorname{Eig}(\lambda_i)$
- (ii) for all $i \in \{2, \dots, n\}$ and $x \in \operatorname{span}(b_i^0, \dots, b_i^{m(\lambda_i)-1})$ we have $\langle Ax, x \rangle_B \leq -b \|x\|_B^2$.

We first obtain that for all vectors $x = \sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k b_i^k$ for some constants (α_i^k) we have $\langle Ax, x \rangle_B \leq -b \|x\|_B^2$. Indeed, since $Ab_i^j \in \operatorname{Eig}(\lambda_i)$ we obtain

$$\begin{aligned} \langle Ax, x \rangle_B &= \left\langle A \cdot \left(\sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k b_i^k \right), \sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k b_i^k \right\rangle_B \\ &= \sum_{i=2}^n \left\langle A \left(\sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k b_i^k \right), \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k b_i^k \right\rangle_B \\ &\quad + \sum_{i=2}^n \sum_{j=2, j \neq i}^n \left\langle \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k A b_i^k, \sum_{k=0}^{m(\lambda_j)-1} \alpha_j^k b_j^k \right\rangle_B \\ &\stackrel{\text{(ii)}}{\leq} -b \sum_{i=2}^n \left\| \left(\sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k b_i^k \right) \right\|_B^2 + 0 \\ &= -b \sum_{i=2}^n \left\langle \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k b_i^k, \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k b_i^k \right\rangle_B \\ &= -b \left\langle \sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k b_i^k, \sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k b_i^k \right\rangle_B \\ &= -b \left\| \sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k b_i^k \right\|_B^2. \end{aligned}$$

As a next step we note that the set $\mathcal{P}(\mathcal{S}) - \bar{m} := \{x \in \mathbb{R}^S : \exists m \in \mathcal{P}(\mathcal{S}) : x = m - \bar{m}\}$ is compact and that

$$(\alpha_i^k)_{i \in \{1, \dots, n\}, k \in \{1, \dots, m(\lambda_i)-1\}} \mapsto \alpha_1^0 \bar{m} + \sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k b_i^k$$

is a homeomorphism. Thus, also

$$P := \left\{ (\alpha_i^k)_{i \in \{1, \dots, n\}, k \in \{1, \dots, m(\lambda_i)-1\}} \in \mathbb{R}^S : \alpha_1^0 \bar{m} + \sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k b_i^k \in \mathcal{P}(\mathcal{S}) - \bar{m} \right\}$$

is compact.

For $x = \alpha_1^0 \bar{m} + \sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k b_i^k \in \mathcal{P}(\mathcal{S}) - \bar{m}$ we moreover obtain that

$$\begin{aligned} & \langle Ax, x \rangle_B \\ &= \left\langle A \left(\alpha_1^0 \bar{m} + \sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k b_i^k \right), \alpha_1^0 \bar{m} + \sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k b_i^k \right\rangle_B \\ &= (\alpha_1^0)^2 \langle A \bar{m}, \bar{m} \rangle_B + \left\langle \alpha_1^0 A \bar{m}, \sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k b_i^k \right\rangle_B + \left\langle \sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k A b_i^k, \alpha_1^0 \bar{m} \right\rangle_B \\ &\quad + \left\langle A \cdot \sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k b_i^k, \sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k b_i^k \right\rangle_B \\ &= 0 + 0 + 0 + \left\langle A \cdot \sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k b_i^k, \sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k b_i^k \right\rangle_B \\ &\leq -b \cdot \left\| \sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k b_i^k \right\|_B^2 \\ &= -b \cdot \frac{\left\| \sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k b_i^k \right\|_B^2}{\underbrace{\left\| \alpha_1^0 \bar{m} + \sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k b_i^k \right\|_B^2}_{=: D(\alpha) > 0}} \cdot \left\| \underbrace{\alpha_1^0 \bar{m} + \sum_{i=2}^n \sum_{k=0}^{m(\lambda_i)-1} \alpha_i^k b_i^k}_{=: x} \right\|_B^2 \\ &\leq -b \underbrace{\left(\min_{\alpha \in P} D(\alpha) \right)}_{=: D} \|x\|_B^2. \end{aligned}$$

Since P is compact and $D(\alpha) > 0$ for all $\alpha \in P$, it immediately follows that $D > 0$.

As a final preparation we note that, by the same reasoning as in the proof of Theorem 3.2, the set $\mathcal{P}(\mathcal{S})$ is flow invariant for $\dot{m}(t) = f^d(m(t))$. Thus, the set $\mathcal{P}(\mathcal{S}) - \bar{m}$ is flow invariant for $\dot{x}(t) = \tilde{f}^d(x(t))$.

By definition of the derivative (which in particular yields $\|\tilde{f}^d(x) - Ax\|_B \in o(\|x\|_B)$ in a neighbourhood of 0) and Cauchy's inequality we have

$$\begin{aligned} 0 &= \lim_{x \rightarrow 0} \frac{\|\tilde{f}^d(x) - Ax\|_B}{\|x\|_B} = \lim_{x \rightarrow 0} \frac{\|\tilde{f}^d(x) - Ax\|_B \cdot \|x\|_B}{\|x\|_B^2} \\ &\geq \lim_{x \rightarrow 0} \frac{\langle \tilde{f}^d(x) - Ax, x \rangle_B}{\|x\|_B^2} = 0. \end{aligned}$$

Since for all $x \in \mathcal{P}(\mathcal{S}) - \bar{m}$ we have $\langle Ax, x \rangle_B \leq -bD\|x\|_B^2$, there is a $\delta > 0$ such that for all $x \in \overline{N_\delta(0)} \cap (\mathcal{P}(\mathcal{S}) - \bar{m})$ it holds that $\langle \tilde{f}^d(x), x \rangle_B \leq -cD\|x\|_B^2$. If $\tilde{x} \in N_\delta(0)$ such that $\mathcal{D}(\tilde{x} + \bar{m}) \neq \{d\}$, then make δ smaller such that $\mathcal{D}(x + \bar{m}) = \{d\}$ for all $x \in \overline{N_\delta(0)} \cap (\mathcal{P}(\mathcal{S}) - \bar{m})$.

Now let $x_0 \in N_\delta(0) \cap (\mathcal{P}(\mathcal{S}) - \bar{m})$. Then by Peano's existence theorem there is a solution $x : [0, \infty) \rightarrow \mathbb{R}^S$ of the initial value problem $\dot{x}(t) = \tilde{f}^d(x(t))$, $x(0) = x_0$ and, furthermore, any solution $x : [0, t_0) \rightarrow \mathbb{R}^S$ can be extended on $[0, \infty)$. Let $x : [0, t_0) \rightarrow \mathbb{R}^S$ be a solution curve of the differential equation $\dot{x}(t) = \tilde{f}^d(x(t))$ in $\overline{N_\delta(0)}$ and assume that $x(t) \neq 0$ for all $0 \leq t \leq t_0$. (If $x(\tilde{t}) = 0$ for some $\tilde{t} \geq 0$, then $x(t) = 0$ for all $t \geq \tilde{t}$.)

Then it holds that

$$\frac{\partial}{\partial t} \|x(t)\|_B = \frac{1}{\|x(t)\|_B} \langle \dot{x}(t), x(t) \rangle_B \leq -cD\|x(t)\|_B, \quad (6)$$

which means that $\|x(t)\|_B$ is strictly decreasing on $[0, t_0]$. Thus, $x(t) \in N_\delta(0)$ for all $t \in [t_0, t_0 + \tilde{\epsilon}]$. Repeating this argument, we obtain by Hirsch and Smale (1974, Section 8.5) that $x(t) \in N_\delta(0)$ for all $t \geq 0$. Furthermore, the estimate (6) yields that $\|x(t)\|_B \leq e^{-cDt} \|x(0)\|_B$, which is the desired exponential convergence. \square

Proof of Theorem 5.1 We first note that if there is a $T \geq 0$ such that the trajectory satisfies $g(m(t)) < 0$ or $g(m(t)) > 0$ for all $t \geq T$, then the solution of (3) is also a solution of $\dot{m}(t) = Q^{d^1}(m(t))^T m(t)$ or $\dot{m}(t) = Q^{d^2}(m(t))^T m(t)$, respectively, which means that $(m(t))_{t \geq T}$ are the marginals of a nonlinear Markov chain. By assumption, we thus obtain convergence towards some stationary point, which is, since $g(m(t)) < 0$ or $g(m(t)) > 0$ for all $t \in [T, \infty)$, a stationary equilibrium.

In case (i) whenever $g(m(T)) = 0$ for some $T \geq 0$, then $g(m(t)) > 0$ for all $t > T$. Indeed, assume that $g(m(t)) = 0$ for all $t \in [T, T + \epsilon]$ for some $\epsilon > 0$. Then there is a measurable function $\lambda : [T, T + \epsilon] \rightarrow [0, 1]$ such that for almost all $t \in [T, T + \epsilon]$ it holds that

$$\frac{\partial}{\partial t} g(m(t)) = \left\langle \lambda(t) Q^{d^1}(m(t))^T m(t) + (1 - \lambda(t)) Q^{d^2}(m(t))^T m(t), \nabla g(m(t)) \right\rangle > 0,$$

which is a contradiction. Similarly, if we assume that $g(m(t)) < 0$ for all $t \in (T, T + \epsilon)$ with $\epsilon > 0$, then it holds for almost all $t \in (T, T + \epsilon)$ that

$$\frac{\partial}{\partial t} g(m(t)) = \left\langle Q^{d^1}(m(t))^T m(t), \nabla g(m(t)) \right\rangle > 0,$$

again a contradiction. Thus, it either holds that $g(m(t)) < 0$ for all $t \geq 0$ or that $g(m(t)) > 0$ for all $t \geq T$ with $T \geq 0$, which by the first observation yields the desired convergence. Analogously, in case (ii) we obtain that whenever $g(m(T)) = 0$ for some $T \geq 0$, then $g(m(t)) < 0$ for all $t > T$.

In case (iii) we have that whenever $g(m(T)) = 0$ for some $T \geq 0$, then $g(m(t)) = 0$ for all $t \geq T$. Indeed, assume that there is a $\tilde{t} > T$ such that $g(m(\tilde{t})) > 0$. Since g is Lipschitz continuous and m is Lipschitz continuous as long as $g(m(\cdot)) > 0$ we have that there is an $\epsilon > 0$ such that $g(m(\tilde{t} - \epsilon)) = 0$ and $g(m(t)) > 0$ for all $\tilde{t} - \epsilon < t \leq \tilde{t}$. In particular for some $\epsilon_2 \in (0, \epsilon_1)$ we have for almost all $\tilde{t} - \epsilon_1 < t \leq \tilde{t} - \epsilon_2$ that

$$0 < \frac{\partial}{\partial t} g(m(t)) = \langle Q^{d^2}(m(t))^T m(t), \nabla g(m(t)) \rangle \leq 0,$$

a contradiction. Similarly, we obtain for the case that there is a $\tilde{t} > T$ such that $g(m(\tilde{t})) < 0$ that there are $0 < \epsilon_2 < \epsilon_1$ such that for almost all $\tilde{t} - \epsilon_1 < t \leq \tilde{t} - \epsilon_2$ we have

$$0 > \frac{\partial}{\partial t} g(m(t)) = \langle Q^{d^1}(m(t))^T m(t), \nabla g(m(t)) \rangle \geq 0,$$

again a contradiction. Thus, either $g(m(t)) < 0$ for all $t \geq 0$, or $g(m(t)) > 0$ for all $t \geq 0$, in which case we obtain convergence towards some stationary equilibrium with a deterministic equilibrium strategy, or there is a $T > 0$ such that $g(m(t)) = 0$ for all $t \geq T$. \square

Lemma B.1 *Let Assumptions A2, A3 and A4 hold. Moreover, assume that for all $d^1, d^2 \in \mathcal{U}$ satisfying $\text{Opt}^{\text{some}}(d^1) \cap \text{Opt}^{\text{some}}(d^2) \neq \emptyset$ we either have*

$$\langle (Q^d(m))^T m, \nabla g_{d^1}(m) \rangle > 0$$

for all $m \in \text{Opt}^{\text{some}}(d^1) \cap \text{Opt}^{\text{some}}(d^2)$ such that $m_i = 0$ for some $i \in \mathcal{S}$ or

$$\langle (Q^d(m))^T m, \nabla g_{d^1}(m) \rangle < 0$$

for all $m \in \text{Opt}^{\text{some}}(d^1) \cap \text{Opt}^{\text{some}}(d^2)$ such that $m_i = 0$ for some $i \in \mathcal{S}$. Then the set $\mathcal{P}(\mathcal{S})$ is flow invariant for $\dot{m} \in F(m)$.

Proof We first note that $\{m \in \mathbb{R}^S : \sum_{i \in \mathcal{S}} m_i = 1\}$ is flow invariant for F since $\sum_{i \in \mathcal{S}} \dot{m}_i(t) = 0$ holds almost surely because all transition rate matrices are conservative. Therefore, if at time \tilde{t}_1 a solution of $\dot{m} \in F(m)$ leaves the set $\mathcal{P}(\mathcal{S})$ at least one component $m_i(\tilde{t}_1)$ has to be zero. By Assumption A4 there are at most two strategies $d^1, d^2 \in \mathcal{U}$ such that $g_{d^1}(m(\tilde{t}_1)) \leq 0$ and $g_{d^2}(m(\tilde{t}_1)) \leq 0$. In particular, there is an $\epsilon > 0$ such that $g_d(\tilde{m}) > 0$ for all $d \neq d^1, d^2$ and $\tilde{m} \in N_\epsilon(m(\tilde{t}_1))$.

Let us define $\phi = g_{d^1}$ and note that $g_{d^2}(m) = g_{d^1}(m)$ for $m \in \text{Opt}^{\text{some}}(d^1) \cap \text{Opt}^{\text{some}}(d^2)$. Thus, the consistency condition yields that one of the following three cases will hold for all $m \in \text{Opt}^{\text{some}}(d^1) \cap \text{Opt}^{\text{some}}(d^2)$ simultaneously:

- $\langle (Q^{d^1}(m))^T m, \nabla \phi(m) \rangle > 0$ and $\langle (Q^{d^2}(m))^T m, \nabla - \phi(m) \rangle < 0$

- $\langle (Q^{d^1}(m))^T m, \nabla \phi(m) \rangle < 0$ and $\langle (Q^{d^2}(m))^T m, \nabla - \phi(m) \rangle > 0$
- $\langle (Q^{d^1}(m))^T m, \nabla \phi(m) \rangle > 0$ and $\langle (Q^{d^2}(m))^T m, \nabla - \phi(m) \rangle > 0$.

In particular, the results presented in Filippov (1988, §4 and §10) yield that the solution in $N_\epsilon(m(\tilde{t}_1))$ is the solution of a classical ordinary differential equation $\dot{m} = (\tilde{Q}(m))^T m$ with $\tilde{Q}(m)$ being $Q^{d^1}(m)$ in the first case, $Q^{d^2}(m)$ in the second case and

$$\frac{\langle (Q^{d^1}(m))^T m, \nabla \phi(m) \rangle}{\langle (Q^{d^1}(m))^T m, \nabla \phi(m) \rangle + (-\langle (Q^{d^2}(m))^T m, \nabla \phi(m) \rangle)} Q^{d^1}(m) \\ + \frac{-\langle (Q^{d^2}(m))^T m, \nabla \phi(m) \rangle}{\langle (Q^{d^1}(m))^T m, \nabla \phi(m) \rangle + (-\langle (Q^{d^2}(m))^T m, \nabla \phi(m) \rangle)} Q^{d^2}(m)$$

in the third case. In all three cases we obtain that $\tilde{Q}(\cdot)$ is Lipschitz continuous. Moreover, we have that $\mathcal{P}(\mathcal{S})$ is flow invariant for $\dot{m} = (\tilde{Q}(m))^T m$ with $\tilde{Q}(m)$: Indeed, it suffices to show that $(\tilde{Q}(m))^T m \in T_{\mathcal{P}(\mathcal{S})}(m)$, which works as in the proof of Theorem 3.2. All in all, this is the desired contradiction. \square

In order to prove Theorem 5.4 we first present two auxiliary lemmata describing flow invariant sets for $\dot{m} \in F(m)$:

Lemma B.2 *Let \leq be an ordering of \mathcal{U} satisfying (iii) and enumerate the deterministic stationary strategies such that $d^1 \leq d^2 \leq \dots \leq d^u$. Furthermore, let $1 \leq k \leq u$. Then the set*

$$\mathcal{P}(\mathcal{S}) \cap \left(\bigcup_{l \geq k} \text{Opt}^{\text{some}}(d^l) \right)$$

is flow invariant for F .

Proof We prove the statement by backward induction on k . Let first $k = u$ and assume that $m : [0, \infty) \rightarrow \mathbb{R}^S$ is a solution of $\dot{m} \in F(m)$ starting in $m_0 \in \mathcal{P}(\mathcal{S}) \cap \text{Opt}^{\text{some}}(d^u)$. By Assumption A5 we immediately have that $m(t) \in \mathcal{P}(\mathcal{S})$ for all $t \geq 0$. Moreover, it is immediate by definition of $\text{Opt}^{\text{some}}(d^u)$ as well as g_{d^u} that $g_{d^u}(m(0)) \leq 0$. We will now prove that whenever there is a $T \geq 0$ such that $g_{d^u}(m(T)) = 0$ then $g_{d^u}(m(t)) \leq 0$ for all $t > T$. This and the continuity of the function g_{d^u} then yields, again by definition of $\text{Opt}^{\text{some}}(d^u)$ as well as g_{d^u} the desired claim.

So assume that there is a $T_1 > 0$ such that $g_{d^u}(m(T_1)) > 0$. Then by the Lipschitz continuity of g_{d^u} there is a $T_2 \geq 0$ and a $\delta > 0$ such that $g_{d^u}(m(T_2)) = 0$ and $g_{d^u}(m(t)) > 0$ for all $t \in (T_2, T_2 + \delta]$. Then by definition of $\text{Opt}^{\text{some}}(d^u)$ as well as g_{d^u} and Assumption A4 we immediately have that there is a unique strategy $d^l \in \mathcal{U} \setminus \{d^u\}$ such that $g_{d^l}(m(T_2)) = 0$.

Moreover, there is an $\epsilon_1 \in (0, \delta)$ such that $g_{d^l}(m(t)) < 0$ for all $t \in (T_2, T_2 + \epsilon_1)$: Indeed, assume that there is no $\epsilon > 0$ satisfying this property. Then for all $n \in \mathbb{N}$

there is a $t_n \in [T, T + \frac{1}{n}]$ such that $g_{d^l}(m(t_n)) \geq 0$. Let n_k^1 be the subsequence such that $g_{d^l}(m(t_{n_k^1})) = 0$ and let n_k^2 be the subsequence such that $g_{d^l}(m(t_{n_k^2})) > 0$. At least one of these sequences consist of infinitely many elements. Let us first assume that $(n_k^1)_{k \in \mathbb{N}}$ consists of infinitely many elements. Then, by definition of $\text{Opt}^{\text{some}}(d^l)$ as well as g_{d^l} , there is another strategy $d^{n_k^1} \neq d^u, d^l$ such that $g_{d^{n_k^1}}(m(t_{n_k^1})) = 0$. Since \mathcal{U} is finite there is at least one strategy d^j that occurs infinite many times in the sequences $(d^{n_k^1})_{k \in \mathbb{N}}$. Let us choose a subsequence $(n_i^3)_{i \in \mathbb{N}}$ such that $d^{n_i^3} = d^j$ for all $i \in \mathbb{N}$. Then we obtain that $(t_{n_i^3})_{i \in \mathbb{N}}$ is a sequence converging to T_2 such that $g_{d^j}(m(t_{n_i^3})) = 0$ for all $i \in \mathbb{N}$. In particular, by continuity, we obtain $g_{d^j}(m(T_2)) = 0$, a contradiction. Let us now assume that $(n_k^2)_{k \in \mathbb{N}}$ consists of infinitely many elements. By Assumption A4 and by definition of g_d there is at least one strategy $d^{n_k^2} \neq d^u, d^l$ such that $g_{d^{n_k^2}}(m(t_{n_k^2})) \leq 0$. Since \mathcal{U} is finite there is at least one strategy d^j that occurs infinitely many times in the sequence $(d^{n_k^2})_{k \in \mathbb{N}}$. Let $(n_k^4)_{k \in \mathbb{N}}$ be the sequence such that $d^j = d^{n_k^4}$. Now we obtain that $(t_{n_k^4})_{k \in \mathbb{N}}$ is a sequence converging to T_2 and that satisfies $g_{d^j}(m(t_{n_k^4})) \leq 0$. Therefore, by continuity of g_{d^j} , we obtain $g_{d^j}(m(T_2)) \leq 0$, a contradiction.

By definition of the ordering \leq , we have that $d^l \leq_{D(d^l)} d^u$ in the acyclic ordering $\leq_{D(d^l)}$ of the graph $D(d^l)$. This means that there is no edge from d^u to d^l in $D(d^l)$, which by definition of the digraph means that

$$\langle (Q^{d^l}(m))^T m, \nabla g_{d^u}(m) \rangle \leq 0$$

for all $m \in \text{Opt}^{\text{some}}(d^u) \cap \text{Opt}^{\text{some}}(d^l)$.

In total, we obtain for some $\epsilon_2 \in (0, \epsilon_1)$ and for almost all $t \in (T_2, T_2 + \epsilon_2)$ that

$$0 < \frac{\partial}{\partial t} g_{d^u}(m(t)) = \langle (Q^{d^l}(m))^T m, \nabla g_{d^u}(m) \rangle < 0$$

holds, a contradiction.

Let now $k < u$ and assume that $m : [0, \infty) \rightarrow \mathbb{R}^S$ is a solution of $\dot{m} \in F(m)$ starting in $m_0 \in \mathcal{P}(\mathcal{S}) \cap \bigcup_{l \geq k} \text{Opt}^{\text{some}}(d^l)$. By Assumption A5 we immediately have that $m(t) \in \mathcal{P}(\mathcal{S})$. Moreover, by the induction hypothesis, the set $\mathcal{P}(\mathcal{S}) \cap \bigcup_{l \geq k+1} \text{Opt}^{\text{some}}(d^l)$ is flow invariant. Therefore, if the trajectory would leave $\mathcal{P}(\mathcal{S}) \cap \bigcup_{l \geq k} \text{Opt}^{\text{some}}(d^l)$, then for $T = \sup\{t \geq 0 : m(t) \in \mathcal{P}(\mathcal{S}) \cap \bigcup_{l \geq k} \text{Opt}^{\text{some}}(d^l)\}$ we would have that $m(T) \in \text{Opt}^{\text{some}}(d^k)$. As before, by definition of $\text{Opt}^{\text{some}}(d^k)$ and g_{d^k} , this is equivalent to $g_{d^k}(m(T)) = 0$ and $g_{d^k}(t) > 0$ for some $\delta > 0$ and all $t \in (T, T + \delta)$.

By definition of $\text{Opt}^{\text{some}}(d)$ and d , Assumption A5 and the requirement that m leaves the set $\mathcal{P}(\mathcal{S}) \cap \bigcup_{l \geq k} \text{Opt}^{\text{some}}(d^l)$ after T we obtain that there is a unique $l < k$ such that $g_{d^l}(m(T)) = 0$. As in the base case we can now prove that there is an $\epsilon_1 \in (0, \delta)$ such that $g_{d^l}(m(t)) < 0$ for all $t \in (T, T + \epsilon_1)$.

The ordering \leq now yields that $d^l \leq_{D(d^l)} d^k$ in the acyclic ordering $\leq_{D(d^l)}$ of the graph $D(d^l)$. This means that there is no edge from d^k to d^l in $D(d^l)$, which by definition of the digraph means that

$$\left\langle (Q^{d^l}(m))^T m, \nabla g_{d^k}(m) \right\rangle \leq 0$$

for all $m \in \text{Opt}^{\text{some}}(d^k) \cap \text{Opt}^{\text{some}}(d^l)$.

In total, we obtain for some $\epsilon_2 \in (0, \epsilon_1)$ and for almost all $t \in (T, T + \epsilon_2)$ that

$$0 < \frac{\partial}{\partial t} g_{d^k}(m(t)) = \left\langle (Q^{d^l}(m))^T m, \nabla g_{d^k}(m) \right\rangle \leq 0$$

holds, a contradiction. \square

Lemma B.3 *Let $d \in \mathcal{U}$ be a deterministic stationary strategy such that for any other strategy $\hat{d} \in \mathcal{U} \setminus \{d\}$ either $\text{Opt}^{\text{some}}(d) \cap \text{Opt}^{\text{some}}(\hat{d}) = \emptyset$ or $\hat{d} \rightarrow d \notin D(\hat{d})$ with*

$$\langle (Q^d(m))^T m, \nabla g_d(m) \rangle < 0 \quad \text{for all } m \in \text{Opt}^{\text{some}}(d) \cap \text{Opt}^{\text{some}}(\hat{d}).$$

Then $\mathcal{P}(\mathcal{S}) \cap \text{Opt}^{\text{unique}}(d)$ is flow invariant for F .

Proof Let $m : [0, \infty) \rightarrow \mathbb{R}^S$ be a solution of $\dot{m} \in F(m)$ starting in $m_0 \in \mathcal{P}(\mathcal{S}) \cap \text{Opt}^{\text{unique}}(d)$. By Assumption A5 it is immediate that $m(t) \in \mathcal{P}(\mathcal{S})$ for all $t \geq 0$. By definition of $\text{Opt}^{\text{unique}}(d)$ it is moreover immediate that $g_d(m(0)) < 0$. Now assume that there is a $T > 0$ such that $g_d(m(T)) = 0$. Then, by definition of $\text{Opt}^{\text{some}}(d)$ and $g_{\hat{d}}$ and Assumption A4, there is a unique strategy $\hat{d} \in \mathcal{U} \setminus \{d\}$ such that $g_{\hat{d}}(m(T)) = 0$. Now the assumption of the lemma yields that

$$\langle (Q^d(m))^T m, \nabla g_d(m) \rangle < 0 \quad \text{for all } m \in \text{Opt}^{\text{some}}(d) \cap \text{Opt}^{\text{some}}(\hat{d}).$$

In total, we obtain, since g_d is twice continuously differentiable, that for some $\epsilon > 0$ and almost all $t \in (T - \epsilon, T)$ we have

$$0 < \frac{\partial}{\partial t} g_d(m(t)) = \langle (Q^d(m(t)))^T m, \nabla g_d(m(t)) \rangle < 0,$$

a contradiction. \square

Proof of Theorem 5.4 We note that if a trajectory stays inside a set $\text{Opt}^{\text{unique}}(d)$ for all $t > T$, then by condition (i) the trajectory will converge towards a stationary point given $Q^d(\cdot)$.

Let d^1, \dots, d^u be an ordering that satisfies (iii) and assume that i is maximal such that $m(0) \in \mathcal{P}(\mathcal{S}) \cap (\bigcup_{k \geq i} \text{Opt}^{\text{some}}(d^k))$, then $m(0) \in \text{Opt}^{\text{some}}(d^i)$. If the trajectory does not leave the set $\mathcal{P}(\mathcal{S}) \cap \text{Opt}^{\text{unique}}(d^i)$ we have by the previous observation convergence towards a deterministic stationary mean field equilibrium. Else, by the flow invariance of $\mathcal{P}(\mathcal{S})$ there is a $t_0 \geq 0$ such that the trajectory will stay in

$$\mathcal{P}(\mathcal{S}) \cap (\text{Opt}^{\text{some}}(d^i) \setminus \text{Opt}^{\text{unique}}(d^i))$$

for all $t \geq t_0$ or there is a $t_1 > 0$ such that $m(t_1) \notin \mathcal{P}(\mathcal{S}) \cap \text{Opt}^{\text{some}}(d^i)$. Then we find, since $\mathcal{P}(\mathcal{S}) \cap (\bigcup_{k \geq i} \text{Opt}^{\text{some}}(d^k))$ is flow invariant an $\hat{i} > i$ such that $m(t_1) \in \mathcal{P}(\mathcal{S}) \cap (\bigcup_{k \geq \hat{i}} \text{Opt}^{\text{some}}(d^k))$ and we can reapply the previous argument.

If we reach the final vertex d of an ordering that satisfies (iii) and there is another strategy \hat{d} that is also the final vertex of an ordering, then we obtain that both $\mathcal{P}(S) \cap \text{Opt}^{\text{some}}(d)$ and $\mathcal{P}(S) \cap \text{Opt}^{\text{some}}(\hat{d})$ are flow invariant, which in particular yields that $\mathcal{P}(S) \cap \text{Opt}^{\text{some}}(d) \cap \text{Opt}^{\text{some}}(\hat{d})$ is flow invariant, which yields that whenever a trajectory leaves $\mathcal{P}(S) \cap \text{Opt}^{\text{unique}}(d)$, then it will remain in

$$\mathcal{P}(S) \cap \text{Opt}^{\text{some}}(d) \cap \text{Opt}^{\text{some}}(\hat{d}) \subseteq \mathcal{P}(S) \cap \text{Opt}^{\text{some}}(d) \setminus \text{Opt}^{\text{unique}}(d)$$

for all times, which proves the claim.

If there is a unique final vertex d then we obtain, by Lemma B.3, that $\mathcal{P}(S) \cap \text{Opt}^{\text{unique}}(d)$ is flow invariant. Thus, if the trajectory leaves the set $\mathcal{P}(S) \cap \text{Opt}^{\text{some}}(d) \setminus \text{Opt}^{\text{unique}}(d)$, then we have convergence towards a deterministic stationary mean field equilibrium. \square

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability Not applicable since no datasets were generated or analysed in the study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Asmussen S (2003) Applied probability and queues, Stochastic modelling and applied probability, vol 51, 2nd edn. Springer, New York, ISBN 0-387-00211-1
- Aubin J-P, Cellina A (1984) Differential inclusions: set-valued maps and viability theory, Grundlehren der mathematischen Wissenschaften, vol 264. Springer, Berlin. <https://doi.org/10.1007/978-3-642-69512-4>
- Bang-Jensen J, Gutin GZ (2010) Digraphs: theory, algorithms and applications. Springer Monographs in Mathematics, 2nd edn. Springer, London, ISBN 978-0-85729-041-0
- Belak C, Hoffmann D, Seifried FT (2021) Continuous-time mean field games with finite state space and common noise. Appl Math Optim 84:3173–3216. <https://doi.org/10.1007/s00245-020-09743-7>
- Bensoussan A, Frehse J, Yam P (2013) Mean field games and mean field type control theory. Springer Briefs in Mathematics. Springer, New York. ISBN 978-1-4614-8507-0
- Besancenot D, Dogguy H (2015) Paradigm shift: a mean field game approach. Bull Econ Res 67(3):289–302. <https://doi.org/10.1111/boer.12024>
- Briani A, Cardaliaguet P (2018) Stable solutions in potential mean field game systems. Nonlinear Differ Equ Appl NoDEA 25(1). <https://doi.org/10.1007/s00030-017-0493-3>
- Cardaliaguet P (2013) Notes on mean field games (from P.-L. Lions' Lectures at Collège de France). <https://www.ceremade.dauphine.fr/~cardaliaguet/MFG20130420.pdf>
- Cardaliaguet P, Hadikhani S (2017) Learning in mean field games: the fictitious play. ESAIM Control Optim Calc Var 23 (2):569–591. <https://doi.org/10.1051/cocv/2016004>

- Carmona R, Delarue F (2018a) Probabilistic theory of mean field games with applications I: mean field FBSDEs, control, and games, Probability Theory and Stochastic Modelling, vol 83. Springer International Publishing. <https://doi.org/10.1007/978-3-319-58920-6>
- Carmona R, Delarue F (2018b) Probabilistic theory of mean field games with applications II: mean field games with common noise and master equations, Probability Theory and Stochastic Modelling, vol 84. Springer International Publishing. <https://doi.org/10.1007/978-3-319-56436-4>
- Carmona R, Wang P (2021) A probabilistic approach to extended finite state mean field games. *Math Oper Res* 46(2):471–502. <https://doi.org/10.1287/moor.2020.1071>
- Cecchin A, Fischer M (2018) Probabilistic approach to finite state mean field games. *Appl Math Optim*. <https://doi.org/10.1007/s00245-018-9488-7>
- Deimling K (1992) Multivalued differential equations, de Gruyter Series in Nonlinear Analysis and Applications, vol 1. Walter de Gruyter & Co., Berlin, ISBN 3-11-013212-5
- Doncel J, Gast N, Gaujal B (2019) Discrete mean field games: existence of equilibria and convergence. *J Dyn Games* 6(3):221–239. <https://doi.org/10.3934/jdg.2019016>
- Filippov AF (1988) Differential equations with discontinuous righthand sides, Mathematics and its Applications, vol 18. Springer Netherlands, Dordrecht. <https://doi.org/10.1007/978-94-015-7793-9>
- Fudenberg D, Levine DK (1998) The theory of learning in games, MIT Press Series on Economic Learning and Social Evolution, vol 2. The MIT Press, Cambridge, ISBN 978-0-262-06194-0
- Gomes DA, Mohr J, Souza RR (2013) Continuous time finite state mean field games. *Appl Math Optim* 68(1):99–143. <https://doi.org/10.1007/s00245-013-9202-8>
- Gomes D, Velho RM, Wolfram M-T (2014) Socio-economic applications of finite state mean field games. *Philos Trans R Soc A* 372:20130405. <https://doi.org/10.1098/rsta.2013.0405>
- Guo X, Hernández-Lerma O (2009) Continuous-time Markov decision processes: theory and applications, Stochastic Modelling and Applied Probability, vol 62. Springer, Berlin, ISBN 978-3-642-26072-8
- Guo X, Hu A, Xu R, Zhang J (2019) Learning mean-field games. In: Wallach H, Larochelle H, Beygelzimer A, d'Alché Buc F, Fox E, Garnett R (eds) *Advances in neural information processing systems*, vol 32
- Hadikhanloo S (2017) Learning in anonymous nonatomic games with applications to first-order mean field games. Preprint. [arXiv:1704.00378](https://arxiv.org/abs/1704.00378)
- Hadikhanloo S (2018) Learning in mean field games. PhD thesis, Université Paris-Dauphine, Paris. http://www.cmap.polytechnique.fr/~saeed.hadikhanloo/PhD_Thesis.pdf
- Hirsch MW, Smale S (1974) Differential equations, dynamical systems, and linear algebra. Pure and Applied Mathematics. A Series of Monographs and Textbooks. Academic Press Inc, Orlando. ISBN 0-12-349550-4
- Hofbauer J, Sandholm WH (2009) Stable games and their dynamics. *J Econ Theory* 144(4):1665–1693. <https://doi.org/10.1016/j.jet.2009.01.007>
- Huang M, Malhamé RP, Caines PE (2006) Large population stochastic dynamic games: closed-loop McKean–Vlasov systems and the Nash certainty equivalence principle. *Commun Inf Syst* 6(3):221–252. <https://projecteuclid.org/euclid.cis/1183728987>
- Kolokoltsov VN (2010) Nonlinear Markov processes and kinetic equations, Cambridge Tracts in Mathematics, vol 182. Cambridge University Press, Cambridge, ISBN 978-0-521-11184-3
- Kolokoltsov VN, Bensoussan A (2016) Mean-field-game model for botnet defense in cyber-security. *Appl Math Optim* 74(3):669–692. <https://doi.org/10.1007/s00245-016-9389-6>
- Kolokoltsov VN, Malafeyev OA (2017) Mean-field-game model of corruption. *Dyn Games Appl* 7(1):34–47. <https://doi.org/10.1007/s13235-015-0175-x>
- Kolokoltsov VN, Malafeyev OA (2018) Corruption and botnet defense: a mean field game approach. *Int J Game Theory* 47:977–999. <https://doi.org/10.1007/s00182-018-0614-1>
- Lasry J-M, Lions P-L (2007) Mean field games. *Jpn J Math* 2(1):229–260. <https://doi.org/10.1007/s11537-007-0657-8>
- Laurière M, Perrin S, Geist M, Pietquin O (2022) Learning mean field games: a survey. [arXiv:2205.12944](https://arxiv.org/abs/2205.12944)
- Logemann H, Ryan EP (2014) Ordinary differential equations: analysis, qualitative theory and control. Springer Undergraduate Mathematics Series. Springer, London. ISBN 978-1-4471-6397-8
- Marden JR (2012) State based potential games. *Automatica* 48(12):3075–3088. <https://doi.org/10.1016/j.automatica.2012.08.037>
- Maskin E, Tirole J (2001) Markov perfect equilibrium: I. Observable actions. *J Econ Theory* 100(2):191–219. <https://doi.org/10.1006/jeth.2000.2785>

- Mouzouni C (2018) On quasi-stationary mean field games models. *Appl Math Optim*. <https://doi.org/10.1007/s00245-018-9484-y>
- Nachbar J (2009) Learning in games. In: Meyers RA (ed) *Encyclopedia of complexity and systems science*. Springer, New York, pp 5177–5187. https://doi.org/10.1007/978-0-387-30440-3_307
- Neumann BA (2019) Stationary equilibria of mean field games with finite state and action space: existence, computation, stability, and a myopic adjustment process. PhD thesis, Universität Hamburg. <https://ediss.sub.uni-hamburg.de/volltexte/2020/10313/>
- Neumann BA (2020) Stationary equilibria of mean field games with finite state and action space. *Dyn Games Appl* 10:845–871. <https://doi.org/10.1007/s13235-019-00345-9>
- Neumann BA (2023) Nonlinear Markov chains with finite state space: invariant distributions and long-term behaviour. *J Appl Probab* 60(1):30–44. <https://doi.org/10.1017/jpr.2022.23>
- Perrin S, Perolat J, Laurière M, Geist M, Elie R, Pietquin O (2020) Fictitious play for mean field games: continuous time analysis and applications. In: Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H (eds) *Advances in neural information processing systems*, vol 33
- Sandholm WH (2010) Population games and evolutionary dynamics. *Economic learning and social evolution*. MIT Press, Cambridge. ISBN 978-0-262-19587-4
- Sandholm WH (2014) Local stability of strict equilibria under evolutionary game dynamics. *J Dyn Games* 1(3):485–495. <https://doi.org/10.3934/jdg.2014.1.485>
- Sandholm WH (2015) Population games and deterministic evolutionary dynamics, chapter 13. In: Young HP, Zamir S (eds) *Handbook of game theory with economic applications*, vol 4. Elsevier, pp 703–778. <https://doi.org/10.1016/B978-0-444-53766-9.00013-6>
- Zusai D (2020) Gains in evolutionary dynamics: a unifying and intuitive approach to linking static and dynamic stability. [arXiv:1805.04898v7](https://arxiv.org/abs/1805.04898v7)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.