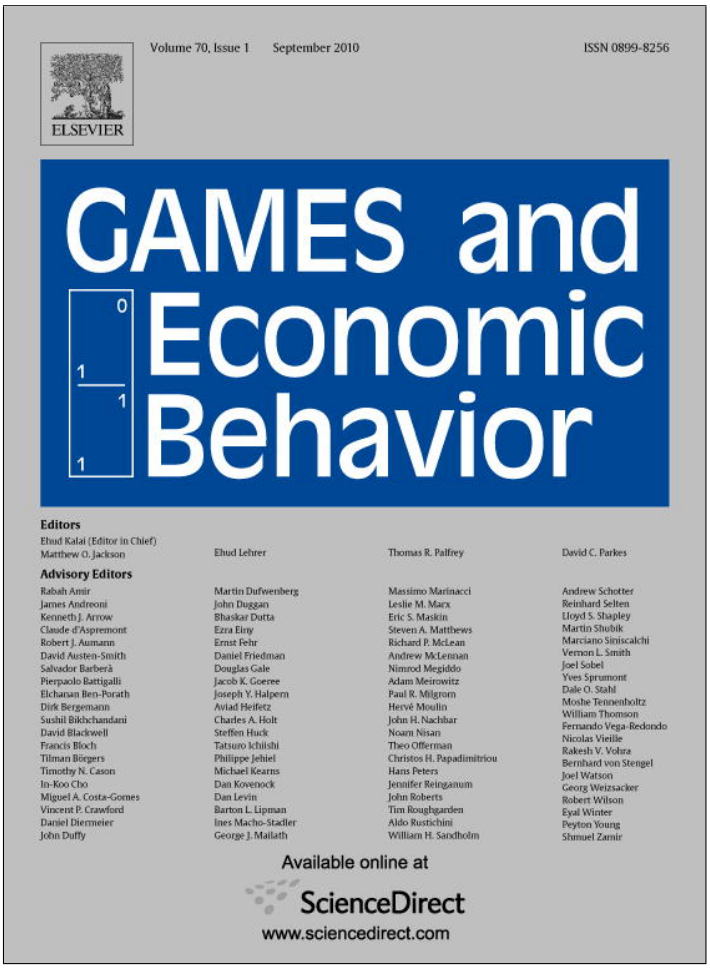

Games and Economic Behavior 70 (2010) 71–83

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

A payoff-based learning procedure and its application to traffic games

Roberto Cominetti [a,1], Emerson Melo [b,*], Sylvain Sorin [c,d]

[a] Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático, Universidad de Chile, Chile
[b] Departamento de Economía, Universidad de Chile and Banco Central de Chile, Chile
[c] Equipe Combinatoire et Optimisation, Faculté de Mathématiques, Université P. et M. Curie – Paris 6, 175 rue du Chevaleret, 75013 Paris, France
[d] Laboratoire d'Econométrie – Ecole Polytechnique, 1 rue Descartes, 75005 Paris, France

ARTICLE INFO

Article history:
Received 6 July 2007
Available online 25 December 2008

Keywords:
Games
Learning
Adaptive dynamics
Stochastic algorithms
Congestion games

ABSTRACT

A stochastic process that describes a payoff-based learning procedure and the associated adaptive behavior of players in a repeated game is considered. The process is shown to converge almost surely towards a stationary state which is characterized as an equilibrium for a related game. The analysis is based on techniques borrowed from the theory of stochastic algorithms and proceeds by studying an associated continuous dynamical system which represents the evolution of the players' evaluations. An application to the case of finitely many users in a congested traffic network with parallel links is considered. Alternative descriptions for the dynamics and the corresponding rest points are discussed, including a Lagrangian representation.

1. Introduction

This paper belongs to the growing literature on the dynamics of repeated games in which players make decisions day after day based on partial information derived from prior experience. The accumulated experience is summarized in a *state variable* that determines the strategic behavior of players through a certain stationary rule. This is the framework of learning and adaptation in games, an area intensively explored in the last decades (Fudenberg and Levine, 1998; Young, 2004). Instead of studying the dynamics at an aggregate level in which the informational and strategic aspects are unspecified, we consider the question from the perspective of the individual player's strategy. At this level the most prominent adaptive procedure is *fictitious play*, early studied by Brown (1951) and Robinson (1951), which assumes that at each stage players choose a best reply to the observed empirical distribution of past moves of their opponents. A recent account of the convergence of this procedure for games with perturbed payoffs can be found in Benaim and Hirsch (1999) introducing tools that we will use here. This variant, called *smooth fictitious play*, is closely tied with Logit random choice and is analyzed in Fudenberg and Levine (1998) and Hofbauer and Sandholm (2002).

The assumption that players are able to record the past moves of their opponents is very stringent for games involving many players with limited observation capacity and bounded rationality. A milder assumption is that each player observes only the outcome vector, namely, the payoff obtained at every stage and the payoff that would have resulted if a different move had been played. Several procedures such as exponential weight (Freund and Schapire, 1999), calibration (Foster and Vohra, 1997), and no-regret procedures *à la Hannan* (Hannan, 1957; Hart, 2005), deal with such limited information contexts: players build statistics of their past performance and infer what the outcome would have been if a different strategy had been played. Eventually, adaptation leads to configurations where no player regrets the choices he makes.

* Corresponding author at: California Institute of Technology, Division of the Humanities and Social Sciences, MC 228-77, Pasadena, CA, USA.
E-mail addresses: rcominet@dim.uchile.cl (R. Cominetti), emelos@hss.caltech.ed (E. Melo), sorin@math.jussieu.fr (S. Sorin).

[1] Supported by FONDAP grant in Applied Mathematics, CONICYT-Chile.

0899-8256/$ – see front matter
doi:10.1016/j.geb.2008.11.012

Although these procedures are flexible and robust, the underlying rationality may still be considered as too demanding in the context of games where players are boundedly rational and less informed. This is the case for traffic in congested networks where a multitude of small players make routing decisions with little or no information about the strategies of the other drivers nor on the actual congestion in the network. The situation is often described as a game where traffic equilibrium is seen as a sort of steady state that emerges from an underlying adaptive mechanism. Wardrop (1952) considered a non-atomic framework ignoring individual drivers and using continuous variables to represent aggregate flows, while Rosenthal (1973) studied the case in which drivers are taken as individual players. Traffic has also been described using random utility models for route choice, leading to the notion of *stochastic user equilibrium* (Daganzo and Sheffi, 1977; Dial, 1971). All these models assume implicitly the existence of a hidden mechanism in travel behavior that leads to equilibrium. Empirical support for this has been given in Avineri and Prashker (2006), Horowitz (1984), Selten et al. (2007) based on laboratory experiments and simulations of different discrete time adaptive dynamics, though it has been observed that the steady states attained may differ from all the standard equilibria and may even depend on the initial conditions. Additional empirical evidence to support the use of discrete choice models in the context of games is presented in McKelvey and Palfrey (1995). From an analytical point of view, the convergence of a class of finite-lag adjustment procedures was established in Cantarella and Cascetta (1995), Cascetta (1989), Davis and Nihan (1993). On a different direction, several continuous time dynamics describing plausible adaptive mechanisms that converge to Wardrop equilibrium were studied in Friesz et al. (1994), Sandholm (2002), Smith (1984), though these models are of an aggregate nature and are not directly linked to the behavior of individual players.

A simpler idea is considered in this paper. We assume that each player has a prior perception or estimate of the payoff performance for each possible move and makes a decision based on this rough information using a random choice rule such as Logit. The payoff of the chosen alternative is then observed and is used to update the perception for that particular move. This procedure is repeated day after day, generating a discrete time stochastic process which we call the *learning process*. The basic ingredients are therefore: a state parameter; a decision rule from states to actions; an updating rule on the state space. This structure is common to many procedures in which the incremental information leads to a change in a state parameter that determines the current behavior through a given stationary map. The specificity here is that the updating rule depends uniquely on the realized payoffs: although players observe only their own payoffs, these values are affected by everybody else's choices revealing information on the game as a whole. The question is whether a simple learning mechanism based on such a minimal piece of information may be sufficient to induce coordination and make the system stabilize to an equilibrium.

It is worth noting that several learning procedures, initially conceived for the case when the complete outcome vector is available, have been adapted to deal with the case where only the realized payoff is known: see Fudenberg and Levine (1998, §4.8) for smooth fictitious play, Auer et al. (2002) for exponential weight, Foster and Vohra (1998) for calibration, and Hart and Mas-Collel (2001) for non-regret. The idea of this kind of approaches is to use the observed payoffs to build an unbiased estimator of the outcome vector, to which the initial version of the procedure is applied. More explicitly, a *pseudo-outcome* vector is defined by the observed payoff divided by the probability with which the actual move was played, on the component corresponding to that move, and completed by zeroes on the other components. Alternatively, the pseudo-outcome vector is built as the empirical average of payoffs obtained on random *exploration stages* having a positive density. The resulting update rules depend not only on the observed payoffs, but also on the probability according to which a move was played as well as on the nature of the stage (exploitation or exploration).

Our process is much simpler in that it relies only on the past sequence of realized moves and payoffs. The idea is closer to the so-called *reinforcement dynamics* in which the only information of a player is her daily payoff (Arthur, 1993; Beggs, 2005; Borgers and Sarin, 1997; Erev and Roth, 1998; Laslier et al., 2001; Posch, 1997), though it differs in the way the state variable is updated as well as in the choice of the decision rule as a function of the state. Usually, reinforcement models use a cumulative rule on a *propensity vector* in which the current payoff is added to the component played, while the remaining components are kept unchanged. A stage-by-stage normalization of the propensity vector leads to a mechanism which is related to the replicator dynamics (Posch, 1997). In this context, convergence has been established for the case of an i.i.d. environment, for zero-sum games, and also for some games with unique equilibria (Laslier et al., 2001; Beggs, 2005). We should also mention here the mechanism proposed in Borgers and Sarin (1997) which uses an averaging rule with payoff dependent weights. Our updating rule uses instead a time average criteria that induces a specific dynamics on perceptions and strategies which appears to be structurally different from the previously studied ones, while preserving the qualitative features of *probabilistic choice* and *sluggish adaptation* (Young, 2004, §2.1).

The paper is organized as follows. Section 2 describes the learning process in the general setting of repeated games, providing sufficient conditions for this process to converge almost surely towards a stationary state which is characterized as an equilibrium for a related game. The analysis relies on techniques borrowed from stochastic algorithms (see e.g. Benaim, 1999; Kushner and Yin, 1997), and proceeds by studying an associated continuous deterministic dynamical system which we call the *adaptive dynamics*. Under suitable assumptions the latter has a unique rest point which is a global attractor, from which the convergence of the learning process follows. In Section 3 we apply the general convergence result to a simple traffic game on a network with parallel links. In this restricted setting the convergence results are established in terms of a "*viscosity parameter*" which represents the amount of noise in players' choices, namely, if noise is large enough the learning process and the associated adaptive dynamics have a unique global attractor. Besides, we obtain a potential function that yields an equivalent Lagrangian description of the dynamics together with alternative characterizations of the rest points.

Finally, we study the case of identical players proving the existence of a unique symmetric mixed equilibrium which is a local attractor for the dynamics under a weaker assumption on the viscosity parameter.

We stress that our model proceeds bottom-up from a simple and explicit behavioral rule to equilibrium: a particular discrete time random model for individual behavior gives rise to an associated continuous time deterministic dynamics which leads ultimately to an equilibrium of a specific associated game. We do not make any claim about the realism of our basic learning rule and several alternatives could be considered. Our contribution is mainly methodological showing that a unified treatment is possible in which a learning process, the adaptive dynamics, and a corresponding notion of equilibrium can be considered in a unified and self-consistent way.

2. Payoff-based adaptive dynamics

2.1. The model

We begin by introducing a dynamical model of adaptive behavior in a repeated game, where each player adjusts iteratively her strategy as a function of past payoffs observed as the game evolves.

Let $\mathcal{P} = \{1, \dots, N\}$ denote the set of players. Each $i \in \mathcal{P}$ is characterized by a finite set S^i of pure strategies and a payoff function $G^i : S^i \times S^{-i} \to \mathbb{R}$ where $S^{-i} = \prod_{j \neq i} S^j$. We denote Δ^i the set of mixed strategies or probability vectors over S^i, and we set $\Delta = \prod_{i \in \mathcal{P}} \Delta^i$. As usual we keep the notation G^i for the multilinear extension of payoffs to the set of mixed strategies.

The game is played repeatedly. At stage n, each player $i \in \mathcal{P}$ selects a move $s_n^i \in S^i$ at random using the mixed strategy

$$\pi_n^i = \sigma^i(x_n^i) \in \Delta^i \tag{1}$$

which depends on a vector $x_n^i = (x_n^{is})_{s \in S^i}$ that represents her perception of the payoff performance of the pure strategies available. Here $\sigma^i : \mathbb{R}^{S^i} \to \Delta^i$ is a continuous map from the space of perceptions to the space of mixed strategies, which describes the stationary behavior rule of player i. We assume throughout that $\sigma^{is}(\cdot)$ is *strictly positive* for all $s \in S^i$.

At the end of the stage, player i observes her own payoff $g_n^i = G^i(s_n^i, s_n^{-i})$, with no additional information about the moves or payoffs of the opponents, and uses this value to adjust her perception of the performance obtained with the pure strategy just played and keeping unchanged the perceptions of the remaining strategies, namely

$$x_{n+1}^{is} = \begin{cases} (1-\gamma_n)x_n^{is} + \gamma_n g_n^i & \text{if } s = s_n^i, \\ x_n^{is} & \text{otherwise,} \end{cases}$$

where $\gamma_n \in (0,1)$ is a sequence of averaging factors with $\sum_n \gamma_n = \infty$ and $\sum_n \gamma_n^2 < \infty$ (a simple choice is $\gamma_n = \frac{1}{n}$). This iteration may be written in vector form as

$$x_{n+1} - x_n = \gamma_n [w_n - x_n] \tag{2}$$

with

$$w_n^{is} = \begin{cases} g_n^i & \text{if } s = s_n^i, \\ x_n^{is} & \text{otherwise.} \end{cases}$$

The distribution of the payoffs and therefore of the random vector w_n is determined by the current perceptions x_n, so that (1) and (2) yield a Markov process for the evolution of perceptions. It may be interpreted as a process in which players simultaneously probe the different pure strategies to *learn* about their payoffs, and adapt their behavior accordingly using the accumulated information to play. The iteration from x_n to x_{n+1} can be decomposed into a chain of elementary steps: the prior perceptions give rise to mixed strategies that lead to moves, which determine the payoffs that are finally used to update the perceptions. Schematically, the procedure just described looks like: $x_n^{is} \rightsquigarrow \pi_n^{is} \rightsquigarrow s_n^i \rightsquigarrow g_n^i \rightsquigarrow x_{n+1}^{is}$. The information gathered at every stage by each player is very limited—only the payoff of the specific move played at that stage—but it conveys implicit information on the behavior of the rest of the players. The basic question we address is whether an iterative procedure based on such a minimal piece of information can lead to coordination among the players on a steady state.

Informally, dividing (2) by the small parameter γ_n the iteration may be interpreted as a finite difference Euler scheme for a related differential equation, except that the right-hand side is not deterministic but a random field. Building on this observation, the theory of stochastic algorithms (see e.g. Benaim, 1999; Benaim et al., 2005) establishes close connections between the asymptotics of the discrete time random process (2) for $n \to \infty$ and the behavior as $t \to \infty$ of the continuous-time deterministic *averaged* dynamics

$$\frac{dx}{dt} = \mathbb{E}(w|x) - x, \tag{3}$$

where $\mathbb{E}(\cdot|x)$ stands for the expectation on the moves induced by the mixed strategies $\sigma^i(x^i)$. In particular, if (3) admits a global attractor x^* (in the sense of dynamical systems) then the discrete process (2) will also converge to x^* with probability

one. This point will be further developed in Section 2.3 using Lyapunov function techniques to establish the existence of such an attractor.

Let us begin by making Eq. (3) more explicit. To this end we consider the space of perceptions $\Omega = \prod_{i\in\mathcal{P}} \mathbb{R}^{S^i}$ and $\Sigma : \Omega \to \Delta$ the profile of mixed strategies of the players at state x

$$\Sigma(x) = \big(\sigma^i(x^i)\big)_{i\in\mathcal{P}}. \tag{4}$$

We introduce the map $C : \Omega \to \Omega$ of expected vector payoffs as a function of the state

$$C(x) = F\big(\Sigma(x)\big) \tag{5}$$

with $F : \Delta \to \Omega$ given by $F(\pi) = (F^i(\pi))_{i\in\mathcal{P}}$, where $F^i(\pi) = (F^{is}(\pi))_{s\in S^i}$ is the expected payoff vector of player i, namely

$$F^{is}(\pi) = G^i\big(s, \pi^{-i}\big). \tag{6}$$

The latter represents the expected payoff for player i when she chooses $s \in S^i$ and the other players use mixed strategies $\{\pi^j\}_{j\neq i}$. Note that F^i does not depend on π^i and that

$$G^i(\pi) = \big\langle \pi^i, F^i(\pi)\big\rangle.$$

Proposition 1. *The continuous dynamics* (3) *may be expressed as*

$$\frac{dx^{is}}{dt} = \sigma^{is}\big(x^i\big)\big[C^{is}(x) - x^{is}\big]. \tag{7}$$

Proof. Denoting $\pi = \Sigma(x)$ and using the definition of the random vector w, the expected value $\mathbb{E}(w|x)$ at x may be computed by conditioning on player i's move as

$$\mathbb{E}\big(w^{is}|x\big) = \pi^{is}G^i\big(s, \pi^{-i}\big) + \big(1 - \pi^{is}\big)x^{is} = \sigma^{is}\big(x^i\big)C^{is}(x) + \big(1 - \sigma^{is}\big(x^i\big)\big)x^{is}$$

which plugged into (3) yields (7). □

We call (7) the *adaptive dynamics* associated to the *learning process* (2). Note that (7) is not directly postulated as a mechanism of adaptive behavior, but instead it is an auxiliary construction to help analyze (2). These dynamics describe the evolution of perceptions and are not stated in the space of mixed strategies as in other adaptive procedures like fictitious play, nor in the space of correlated strategies such as for no-regret or reinforcement dynamics. Moreover, we recall that fictitious play requires the knowledge of all the past moves of the opponents, while no-regret procedures rely on the knowledge of $G^i(a, s^{-i})$ for all $a \in S^i$. Concerning the cumulative proportional reinforcement rule, the state variable is a vector indexed over the set of pure strategies where each component is the sum of the payoffs obtained when that pure strategy was played, while the decision rule is the normalized state variable, so it corresponds to a different dynamics.

2.2. Rest points and perturbed game

According to the general results on stochastic algorithms, the rest points of the continuous dynamics (7) are natural candidates to be limit points for the stochastic process (2). Since $\sigma^{is}(x^i) > 0$ these rest points are the fixed points of the map $x \mapsto C(x)$ whose existence follows easily from Brouwer's theorem if one notes that this map is continuous with bounded range. In this subsection we describe the nature of these rest points focusing on the case where the profile map $\Sigma(x)$ is given by a Logit discrete choice model.

Let $\mathcal{E}$ denote the set of rest points for (7). We note that a point $x \in \mathcal{E}$ is completely characterized by its image $\pi = \Sigma(x)$. To see this it suffices to restate the fixed point equation $x = C(x)$ as a coupled system in (x, π)

$$\begin{cases} \pi = \Sigma(x), \\ x = F(\pi) \end{cases} \tag{8}$$

so that for $x \in \mathcal{E}$ the map $x \mapsto \Sigma(x)$ has an inverse given by $\pi \mapsto F(\pi)$. We state this observation in the following

Proposition 2. *The map $x \mapsto \Sigma(x)$ is one-to-one over the set $\mathcal{E}$.*

A particular choice for the map $\sigma^i(\cdot)$ is given by the Logit rule

$$\sigma^{is}\big(x^i\big) = \frac{\exp(\beta_i x^{is})}{\sum_{a\in S^i}\exp(\beta_i x^{ia})} \tag{9}$$

where the parameter $\beta_i > 0$ has a smoothing effect with $\beta_i \downarrow 0$ leading to a uniform choice while for $\beta_i \uparrow \infty$ the probability concentrates on the pure strategies with higher perceptions. We notice the formal similarity with smooth fictitious play

where x^{is} is replaced by the average past payoff that would have been produced by the move s (which is unknown in the current framework). A closer connection may be established with the notion of *quantal response equilibria* introduced in McKelvey and Palfrey (1995) which are exactly the π's corresponding to system (8), that is to say, the solutions of $\pi = \Sigma(F(\pi))$. As a matter of fact, under the Logit choice formula the one-to-one correspondence between rest points x and the associated π's allows to establish a link between the set $\mathcal{E}$ and the Nash equilibria for a related N-person game $\mathcal{G}$ defined by strategy sets Δ^i for $i \in \mathcal{P}$ and payoff functions $\mathcal{G}^i : \Delta^i \times \Delta^{-i} \to \mathbb{R}$ given by

$$\mathcal{G}^i(\pi) = \langle \pi^i, F^i(\pi) \rangle - \frac{1}{\beta_i} \sum_{s \in S^i} \pi^{is} \big[\ln \pi^{is} - 1\big]$$

which is a perturbation of the original game by an entropy term.

Proposition 3. *If the maps $\sigma^i(\cdot)$ are given by the Logit rule* (9) *then $\Sigma(\mathcal{E})$ is the set of Nash equilibria of the perturbed game $\mathcal{G}$.*

Proof. A well-known characterization of the Logit probabilities gives

$$\sigma^i(x^i) = \arg\max_{\pi^i \in \Delta^i} \langle \pi^i, x^i \rangle - \frac{1}{\beta_i} \sum_{s \in S^i} \pi^{is} \big[\ln \pi^{is} - 1\big].$$

Setting $x^i = F^i(\pi)$ and since this expression does not depend on π^i, Nash equilibria of $\mathcal{G}$ are characterized by $\pi^i = \sigma^i(x^i)$ with $x^i = F^i(\pi)$. But this is precisely (8) so that $\Sigma(\mathcal{E})$ is the set of Nash equilibria of $\mathcal{G}$. □

Remark. The previous characterization extends to the case where the maps $\sigma^i(\cdot)$ are given by more general discrete choice models, namely, player i selects an element $s \in S^i$ that yields a maximal utility $x^{is} + \varepsilon^{is}$ where ε^{is} are non-atomic random variables (the Logit model corresponds to the case when $\{\varepsilon^{is}\}_{s \in S^i}$ are independent Gumbel variables with shape parameter β_i). In this more general framework the probabilities

$$\sigma^{is}(x^i) = \mathbb{P}\big(s \text{ maximizes } x^{is} + \varepsilon^{is}\big)$$

can be expressed as $\sigma^i(x^i) = \nabla \varphi^i(x^i)$ with $\varphi^i(x^i) = \mathbb{E}[\max_{s \in S^i}\{x^{is} + \varepsilon^{is}\}]$ which is smooth and convex. The perturbed payoff functions are now given by

$$\mathcal{G}^i(\pi) = \langle \pi^i, F^i(\pi) \rangle - \theta^i(\pi^i)$$

with $\theta^i(\cdot)$ the Fenchel conjugate of the convex function $\varphi^i(\cdot)$.

2.3. Asymptotic convergence of the dynamics

As mentioned earlier, the asymptotics of (2) and (7) are intimately linked. More precisely, since payoffs are bounded the same holds for any sequence x_n generated by (2), and therefore combining Benaim (1999, Propositions 4.1 and 4.2) and the Limit Set Theorem (Benaim, 1999, Theorem 5.7) it follows that the set of accumulation points of the discrete time random process x_n is almost surely an *internally chain transitive set* (ICT) for the deterministic continuous time dynamics (7). The latter is a strong notion of invariant set for dynamical systems, which allows for the possibility of introducing asymptotically vanishing shocks in the dynamics. For the precise definitions and results, which are somewhat technical, we refer to Benaim (1999). For our purposes it suffices to mention that, according to Benaim (1999, Corollary 5.4), if (7) has a unique rest point $\bar{x}$ which is a global attractor then it is the only ICT and x_n converges to $\bar{x}$ almost surely.

Theorem 4. *If $C : \Omega \to \Omega$ is a $\|\cdot\|_\infty$-contraction then its unique fixed point $\bar{x} \in \Omega$ is a global attractor for the adaptive dynamics* (7) *and the learning process* (2) *converges almost surely towards $\bar{x}$.*

Proof. Let $\ell \in [0, 1)$ be a Lipschitz constant for $C(\cdot)$. The existence and uniqueness of $\bar{x}$ is clear, while almost sure convergence of (2) will follow from the previously cited results in stochastic approximation together with (Benaim, 1999, Corollary 6.6), provided that we exhibit a strict Lyapunov function with a unique minimum at $\bar{x}$. We claim that $\Phi(x) = \|x - \bar{x}\|_\infty$ has this property.

Since $\Phi(x(t))$ is the maximum of the smooth functions $\pm(x^{is}(t) - \bar{x}^{is})$, it is absolutely continuous and its derivative coincides with the derivative of one the functions attaining the max. Specifically, let $i \in \mathcal{P}$ and $s \in S^i$ be such that $\Phi(x(t)) = |x^{is}(t) - \bar{x}^{is}|$. If $x^{is}(t) \geqslant \bar{x}^{is}$, using the fixed point property $\bar{x}^{is} = C^{is}(\bar{x})$ we get

$$\frac{d}{dt}\big[x^{is}(t) - \bar{x}^{is}\big] = \sigma^{is}(x^i)\big[C^{is}(x) - C^{is}(\bar{x}) + \bar{x}^{is} - x^{is}\big] \leqslant \sigma^{is}(x^i)\big[\ell \|x - \bar{x}\|_\infty + \bar{x}^{is} - x^{is}\big] = -\sigma^{is}(x^i)[1 - \ell]\Phi(x)$$

and a similar argument for the case $x^{is}(t) < \bar{x}^{is}$ then yields

$$\frac{d}{dt}\Phi\big(x(t)\big) \leqslant -\min_{is} \sigma^{is}\big(x^i(t)\big)[1 - \ell]\Phi\big(x(t)\big).$$

Now, since $C^{is}(x)$ is bounded it follows from (7) that the same holds for $x(t)$ and then $\sigma^{is}(x^i(t))$ stays bounded away from 0 so that $\frac{d}{dt}\Phi(x(t)) \leqslant -\varepsilon\,\Phi(x(t))$ for some $\varepsilon > 0$. This implies that Φ is a Lyapunov function which decreases to 0 exponentially fast along the trajectories of (7), and since $\bar{x}$ is the unique point with $\Phi(\bar{x}) = 0$ the conclusion follows. □

It is worth noting that the convergence of the state variables $x_n \to \bar{x}$ for the learning dynamics (2), entails the convergence of the corresponding mixed strategies $\pi_n^i = \sigma^i(x_n^i)$ and therefore of the behavior of players.

An explicit condition for $C(\cdot)$ to be a contraction is obtained as follows. Let $\omega = \max_{i\in\mathcal{P}} \sum_{j\neq i} \beta_j$ and take θ an upper bound for the impact over a player's payoff when a single player changes her move, namely

$$\left|G^i(s,u) - G^i(s,v)\right| \leqslant \theta$$

for each player $i \in \mathcal{P}$, every pure strategy $s \in S^i$, and all pairs $u, v \in S^{-i}$ such that $u^j = v^j$ except for one $j \neq i$.

Proposition 5. *Under the Logit rule* (9)*, if* $2\omega\theta < 1$ *then* $C(\cdot)$ *is a* $\|\cdot\|_\infty$*-contraction.*

Proof. Consider the difference $C^{is}(x) - C^{is}(y) = F^{is}(\Sigma(x)) - F^{is}(\Sigma(y))$ for a fixed player $i \in \mathcal{P}$ and pure strategy $s \in S^i$. We may write this difference as a telescopic sum of the terms $\Lambda_j = F^{is}(\pi_j) - F^{is}(\pi_{j-1})$ for $j = 1, \ldots, N$ where

$$\pi_j = \left(\sigma^1(x^1), \ldots, \sigma^j(x^j), \sigma^{j+1}(y^{j+1}), \ldots, \sigma^N(y^N)\right).$$

Since $F^{is}(\pi)$ does not depend on π^i we have $\Lambda_i = 0$. For the remaining terms we note that they can be expressed as $\Lambda_j = \langle A_j, \sigma^j(x^j) - \sigma^j(y^j)\rangle$ where for $t \in S^j$ we put

$$A_j^t = \sum_{u\in S^{-i},\ u^j = t} G^i(s,u) \prod_{k\neq i,j} \pi_j^{ku^k}.$$

Moreover, since both $\sigma^j(x^j)$ and $\sigma^j(y^j)$ belong to the unit simplex Δ^j, we may also write $\Lambda_j = \langle A_j - A_j^r\mathbb{1}, \sigma^j(x^j) - \sigma^j(y^j)\rangle$ for any fixed $r \in S^j$. It is now easy to see that $|A_j^t - A_j^r| \leqslant \theta$ and therefore we deduce

$$\left|C^{is}(x) - C^{is}(y)\right| \leqslant \theta \sum_{j\neq i} \left\|\sigma^j(x^j) - \sigma^j(y^j)\right\|_1. \tag{10}$$

Take $w_j \in \mathbb{R}^{S^j}$ with $\|w_j\|_\infty = 1$ and $\|\sigma^j(x^j) - \sigma^j(y^j)\|_1 = \langle w_j, \sigma^j(x^j) - \sigma^j(y^j)\rangle$, so that using the mean value theorem we may find $z^j \in [x^j, y^j]$ with

$$\left\|\sigma^j(x^j) - \sigma^j(y^j)\right\|_1 = \sum_{t\in S^j} w_j^t \left\langle \nabla\sigma^{jt}(z^j), x^j - y^j\right\rangle \leqslant \sum_{t\in S^j} \left\|\nabla\sigma^{jt}(z^j)\right\|_1 \left\|x^j - y^j\right\|_\infty.$$

Using Lemma 6 below, together with the fact that $\sigma^j(z^j) \in \Delta^j$, we get

$$\left\|\sigma^j(x^j) - \sigma^j(y^j)\right\|_1 \leqslant \sum_{t\in S^j} 2\beta_j\sigma^{jt}(z^j)\left\|x^j - y^j\right\|_\infty \leqslant 2\beta_j\|x - y\|_\infty$$

which combined with (10) gives finally

$$\left|C^{is}(x) - C^{is}(y)\right| \leqslant 2\omega\theta\|x - y\|_\infty. \qquad \square$$

Lemma 6. *For each* $i \in \mathcal{P}$ *and* $s \in S^i$ *the Logit rule* (9) *satisfies*

$$\left\|\nabla\sigma^{is}(x^i)\right\|_1 = 2\beta_i\sigma^{is}(x^i)\left(1 - \sigma^{is}(x^i)\right).$$

Proof. Let $\pi^i = \sigma^i(x^i)$. A direct computation gives $\frac{\partial\sigma^{is}}{\partial x^{it}} = \beta_i\pi^{is}(\delta_{st} - \pi^{it})$ with $\delta_{st} = 1$ if $s = t$ and $\delta_{st} = 0$ otherwise, from which we get

$$\left\|\nabla\sigma^{is}(x^i)\right\|_1 = \beta_i\pi^{is}\sum_{t\in S^i}\left|\delta_{st} - \pi^{it}\right| = 2\beta_i\pi^{is}\left(1 - \pi^{is}\right). \qquad \square$$

3. An application to traffic games

In this section we use the previous framework to model the adaptive behavior of drivers in a congested traffic network. The setting for the *traffic game* is as follows. Each day a set of N users, $i \in \mathcal{P}$, choose one among M alternative routes from a set $\mathcal{R}$. The combined choices of all players determine the total route loads and the corresponding travel times. Each user experiences only the cost of the route chosen on that day and uses this information to adjust the perception for that particular route, affecting the mixed strategy to be played in the next stage.

More precisely, a route $r \in \mathcal{R}$ is characterized by an increasing sequence $c_1^r \leqslant \cdots \leqslant c_N^r$ where c_u^r represents the average travel time of the route when it carries a load of u users. The set of pure strategies for each player $i \in \mathcal{P}$ is $S^i = \mathcal{R}$ and if $r_n^j \in \mathcal{R}$ denotes the route chosen by each player j at stage n, then the payoff to player i is given as the negative of the experienced travel time $g_n^i = G^i(r_n) = -c_u^r$ with $r = r_n^i$ and $u = \#\{j \in \mathcal{P}:\ r_n^j = r\}$.

We assume that the route r_n^i is randomly chosen according to a mixed strategy $\pi_n^i = \sigma^i(x_n^i)$ which depends on prior perceptions about route payoffs through a Logit model

$$\sigma^{ir}(x^i) = \frac{\exp(\beta_i x^{ir})}{\sum_{a\in\mathcal{R}} \exp(\beta_i x^{ia})}, \tag{11}$$

while the evolution of perceptions is governed by (2) as in Section 2.

In the model we assume that all users on route r experience exactly the same travel time c_u^r, though the analysis remains unchanged if we merely suppose that each $i \in \mathcal{P}$ observes a random time $\tilde{c}^{ir}$ with conditional expected value c_u^r given the number u of users that choose r. On the other hand, the network topology here is very simple with only a set of parallel routes, and a natural extension is to consider more general networks. Note however that the parallel link structure allows to model more complex decision problems such as the simultaneous choice of route and departure time, by using the standard trick of replacing a physical route by a set of parallel links that represent the route at different time windows.

3.1. Potential function and global attractor

In the traffic game setting, the vector payoff map $F(\cdot)$ defined by (6) can be expressed as the gradient of a potential function[2] which is inspired from Rosenthal (1973). Namely, consider the map $H : [0,1]^{\mathcal{P}\times\mathcal{R}} \to \mathbb{R}$ defined by

$$H(\pi) = -\mathbb{E}_\pi^B\left[\sum_{r\in\mathcal{R}}\sum_{u=1}^{U^r} c_u^r\right], \tag{12}$$

where $\mathbb{E}_\pi^B$ denotes the expectation with respect to the random variables $U^r = \sum_{i\in\mathcal{P}} X^{ir}$ with X^{ir} independent non-homogeneous Bernoulli random variables such that $\mathbb{P}(X^{ir} = 1) = \pi^{ir}$.

A relevant technical remark is in order here. We observe that $H(\pi)$ was defined for $\pi \in [0,1]^{\mathcal{P}\times\mathcal{R}}$ and not only for $\pi \in \Delta$, which allows to differentiate H with respect to each variable π^{ir} independently, ignoring the constraints $\sum_{r\in\mathcal{R}} \pi^{ir} = 1$. As a result of this independence assumption, the random variable X^{ir} *cannot* be identified with the indicator Y^{ir} of the event *"player i chooses route r"* for which we do have these constraints: each player $i \in \mathcal{P}$ must choose one and only one route $r \in \mathcal{R}$ so that the family $\{Y^{ir}\}_{r\in\mathcal{R}}$ is not independent. However, since player i chooses his route independently from other players, once a route $r \in \mathcal{R}$ is *fixed* we do have that $\{Y^{kr}\}_{k\neq i}$ is an independent family, so that the distinction between X^{kr} and Y^{kr} becomes superfluous when computing expected payoffs as shown next.

Lemma 7. *For any given $r \in \mathcal{R}$ and $i \in \mathcal{P}$ let $U_i^r = \sum_{k\neq i} X^{kr}$ with X^{kr} independent non-homogeneous Bernoulli's such that $\mathbb{P}(X^{kr} = 1) = \pi^{kr}$. Then*

$$F^{ir}(\pi) = \mathbb{E}_\pi^B\left[-c_{U^r}^r \mid X^{ir} = 1\right] = \mathbb{E}_\pi^B\left[-c_{U_i^r+1}^r\right]. \tag{13}$$

Proof. By definition we have $F^{ir}(\pi) = \mathbb{E}[-c_{V_i^r+1}^r]$ with the expectation taken with respect to the random variable $V_i^r = \sum_{k\neq i} Y^{kr}$ where Y^{kr} denotes the indicator of the event *"player k chooses route r."* Since r is fixed, the variables $\{Y^{kr}\}_{k\neq i}$ are independent Bernoulli's with $\mathbb{P}(Y^{kr} = 1) = \pi^{kr}$ so we may replace them by X^{kr}, that is to say

$$F^{ir}(\pi) = \mathbb{E}_\pi^B\left[-c_{U_i^r+1}^r\right] = \mathbb{E}_\pi^B\left[-c_{U^r}^r \mid X^{ir} = 1\right]. \qquad \square$$

We deduce from this property that H is a potential.

Proposition 8. *$F(\pi) = \nabla H(\pi)$ for all $\pi \in \Delta$.*

Proof. We note that $H(\pi) = -\sum_{r\in\mathcal{R}} \mathbb{E}_\pi^B[\sum_{u=1}^{U^r} c_u^r]$ so that conditioning on the variables $\{X^{ir}\}_{r\in\mathcal{R}}$ we get

$$H(\pi) = -\sum_{r\in\mathcal{R}}\left[\pi^{ir}\mathbb{E}_\pi^B\left(\sum_{u=1}^{U_i^r+1} c_u^r\right) + (1-\pi^{ir})\mathbb{E}_\pi^B\left(\sum_{u=1}^{U_i^r} c_u^r\right)\right]$$

[2] Although the traffic game is also a *potential game* in the sense of Monderer and Shapley (1996), our notion of potential is closer to the one in Sandholm (2001) in the context of games with a continuous set of players and which simply means that the vector payoff is obtained as the gradient of a real valued smooth function.

which combined with (13) yields

$$\frac{\partial H}{\partial \pi^{ir}}(\pi) = \mathbb{E}^B_\pi\left[\sum_{u=1}^{U_i^r} c_u^r\right] - \mathbb{E}^B_\pi\left[\sum_{u=1}^{U_i^r+1} c_u^r\right] = \mathbb{E}^B_\pi\left[-c^r_{U_i^r+1}\right] = F^{ir}(\pi). \qquad \square$$

Using formula (13), or equivalently Proposition 8, we may obtain the Lipschitz estimates required to study the convergence of the learning process. In particular we note that $F^{ir}(\pi)$ turns out to be a symmetric polynomial in the variables $(\pi^{kr})_{k\neq i}$ only, and does not depend on the probabilities with which the users choose other routes. This allows us to obtain sufficient conditions which are tighter than the one derived from Proposition 5. The following results are expressed in terms of a parameter that measures the congestion effect produced by an additional user, namely

$$\delta = \max\left\{c_u^r - c_{u-1}^r\colon\ r\in\mathcal{R};\ u = 2,\dots,N\right\}. \tag{14}$$

Lemma 9. *The second derivatives of H are all zero except for*

$$\frac{\partial^2 H}{\partial \pi^{jr}\partial \pi^{ir}}(\pi) = \mathbb{E}^B_\pi\left[c^r_{U^r_{ij}+1} - c^r_{U^r_{ij}+2}\right] \in [-\delta, 0] \tag{15}$$

with $j\neq i$, where $U^r_{ij} = \sum_{k\neq i,j} X^{kr}$.

Proof. We just noted that $\frac{\partial H}{\partial \pi^{ir}}(\pi) = \mathbb{E}^B_\pi[-c^r_{U_i^r+1}]$ depends only on $(\pi^{kr})_{k\neq i}$. Also, conditioning on X^{jr} we get

$$\frac{\partial H}{\partial \pi^{ir}}(\pi) = \pi^{jr}\mathbb{E}^B_\pi\left[-c^r_{U^r_{ij}+2}\right] + \left(1-\pi^{jr}\right)\mathbb{E}^B_\pi\left[-c^r_{U^r_{ij}+1}\right]$$

from which (15) follows at once. $\square$

As a corollary of these estimates and Theorem 4 we obtain the following global convergence result. Recall that $\omega = \max_{i\in\mathcal{P}}\sum_{j\neq i}\beta_j$.

Theorem 10. *Assume in the traffic game that $\omega\delta < 2$. Then the corresponding adaptive dynamics* (7) *has a unique rest point $\bar{x}$ which is a global attractor and the process* (2) *converges almost surely to $\bar{x}$.*

Proof. Proposition 8 gives $C^{ir}(x) = F^{ir}(\Sigma(x)) = \frac{\partial H}{\partial \pi^{ir}}(\Sigma(x))$, and since $\frac{\partial H}{\partial \pi^{ir}}(\pi)$ depends only on $\{\pi^{kr}\}_{k\neq i}$, using Lemma 9 we deduce

$$\left|C^{ir}(x) - C^{ir}(y)\right| = \left|\frac{\partial H}{\partial \pi^{ir}}\left(\Sigma(x)\right) - \frac{\partial H}{\partial \pi^{ir}}\left(\Sigma(y)\right)\right| \leqslant \delta\sum_{j\neq i}\left|\sigma^{jr}\left(x^j\right) - \sigma^{jr}\left(y^j\right)\right|.$$

Now Lemma 6 and the inequality $\sigma(1-\sigma)\leqslant\frac{1}{4}$ gives us

$$\left|\sigma^{jr}\left(x^j\right) - \sigma^{jr}\left(y^j\right)\right| \leqslant \frac{1}{2}\beta_j\left\|x^j - y^j\right\|_\infty \leqslant \frac{1}{2}\beta_j\|x-y\|_\infty$$

which combined with the previous estimate yields

$$\left|C^{ir}(x) - C^{ir}(y)\right| \leqslant \frac{1}{2}\delta\sum_{j\neq i}\beta_j\|x-y\|_\infty \leqslant \frac{1}{2}\omega\delta\|x-y\|_\infty.$$

Thus $C(\cdot)$ is a $\|\cdot\|_\infty$-contraction and we may conclude using Theorem 4. $\square$

To interpret this result we note that the β_i's in the Logit formula are inversely proportional to the standard deviation of the random terms in the discrete choice model. Thus, the condition $\omega\delta < 2$ requires either a weak congestion effect (small δ) or a sufficiently large noise (small ω). Although this condition is sharper than the one obtained in Proposition 5, the parameter ω involves sums of β_i's so that it becomes more and more stringent as the number of players increases. In the sequel we show that uniqueness of the rest point still holds under the much weaker condition $\beta_i\delta < 1$ for all $i\in\mathcal{P}$, and even for $\beta_i\delta < 2$ in the case of linear costs ($c_u^r = a^r + \delta^r u$) or when players are symmetric ($\beta_i\equiv\beta$).

It is important to note that at lower noise levels (large β_i's) player behavior becomes increasingly deterministic and multiple pure equilibria will coexist, as in the case of the market entry game proposed in Selten and Guth (1982) which in our setting corresponds to the special case of 2 roads with linear costs and symmetric players. For this game, two alternative learning dynamics were analyzed in Duffy and Hopkins (2005): a proportional reinforcement rule and a Logit rule based on hypothetical reinforcement (which requires further information about the opponents' moves). In the first case convergence to a pure Nash equilibrium is established, while in the second the analysis is done at small noise levels proving convergence towards a perturbed pure Nash equilibrium.

3.2. Lagrangian description of the dynamics

The potential function $H(\cdot)$ allows to rewrite the dynamics (7) in several alternative forms. A straightforward substitution yields

$$\dot{x}^{ir} = \sigma^{ir}(x^i)\left[\frac{\partial H}{\partial \pi^{ir}}\big(\Sigma(x)\big) - x^{ir}\right] \tag{16}$$

so that defining

$$\Psi(\pi) = H(\pi) - \sum_{i\in\mathcal{P}} \frac{1}{\beta_i} \sum_{r\in\mathcal{R}} \pi^{ir}\big[\ln(\pi^{ir}) - 1\big],$$
$$\lambda^i(x^i) = \frac{1}{\beta_i} \ln\Big(\sum_{r\in\mathcal{R}} \exp(\beta_i x^{ir})\Big)$$

one may also put it as

$$\dot{x}^{ir} = \sigma^{ir}(x^i)\left[\frac{\partial \Psi}{\partial \pi^{ir}}\big(\Sigma(x)\big) - \lambda^i(x^i)\right]. \tag{17}$$

Now, setting $\mu^i = \lambda^i(x^i)$ we have $\sigma^{ir}(x^i) = \bar{\pi}^{ir}(x^{ir}, \mu^i) \triangleq \exp[\beta_i(x^{ir} - \mu^i)]$. If instead of considering μ^i as a function of x^i we treat it as an independent variable we find $\frac{\partial \bar{\pi}^{ir}}{\partial x^{ir}} = \beta_i \bar{\pi}^{ir}$, and then introducing the Lagrangians

$$\mathcal{L}(\pi;\mu) = \Psi(\pi) - \sum_{i\in\mathcal{P}} \mu^i\Big[\sum_{r\in\mathcal{R}} \pi^{ir} - 1\Big],$$
$$L(x;\mu) = \mathcal{L}\big(\bar{\pi}(x,\mu);\mu\big)$$

we may rewrite the adaptive dynamics in gradient form

$$\dot{x}^{ir} = \frac{1}{\beta_i}\frac{\partial L}{\partial x^{ir}}\big(x;\lambda(x)\big). \tag{18}$$

Alternatively we may differentiate $\mu^i = \lambda^i(x^i)$ in order to get

$$\dot{\mu}^i = \sum_{r\in\mathcal{R}} \bar{\pi}^{ir}(x^{ir},\mu^i)\dot{x}^{ir} = \frac{1}{\beta_i}\sum_{r\in\mathcal{R}} \bar{\pi}^{ir}\big(x^{ir},\mu^i\big)\frac{\partial L}{\partial x^{ir}}(x;\mu)$$

which may be integrated back to yield $\mu^i = \lambda^i(x^i)$ as unique solution, so that (18) is also equivalent to the system of coupled differential equations

$$\begin{cases} \dot{x}^{ir} = \dfrac{1}{\beta_i}\dfrac{\partial L}{\partial x^{ir}}(x;\mu), \\ \dot{\mu}^i = \dfrac{1}{\beta_i}\displaystyle\sum_{r\in\mathcal{R}} \bar{\pi}^{ir}\big(x^{ir},\mu^i\big)\frac{\partial L}{\partial x^{ir}}(x;\mu). \end{cases} \tag{19}$$

Finally, all these dynamics may also be expressed in terms of the evolution of the probabilities π^{ir} as

$$\begin{cases} \dot{\pi}^{ir} = \beta_i \pi^{ir}\left[\pi^{ir}\dfrac{\partial \mathcal{L}}{\partial \pi^{ir}}(\pi;\mu) - \dot{\mu}^i\right], \\ \dot{\mu}^i = \displaystyle\sum_{r\in\mathcal{R}} \big(\pi^{ir}\big)^2 \frac{\partial \mathcal{L}}{\partial \pi^{ir}}(\pi;\mu). \end{cases} \tag{20}$$

We stress that (16)–(20) are equivalent ways to describe the adaptive dynamics (7), so they provide alternative means for studying the convergence of the learning process. In particular, (17) may be interpreted as a gradient flow for finding critical points of the functional Ψ on the product of the unit simplices defined by $\sum_{r\in\mathcal{R}} \pi^{ir} = 1$ (even if the dynamics are in the x-space), while (20) can be seen as a dynamical system that searches for saddle points of the Lagrangian $\mathcal{L}$ with the variables μ^i playing the role of multipliers. We show next that these critical points are closely related to the rest points of the adaptive dynamics.

Proposition 11. *Let $x \in \Omega$ and $\pi = \Sigma(x)$. The following are equivalent*:

(a) $x \in \mathcal{E}$,
(b) $\nabla_x L(x,\mu) = 0$ *for* $\mu = \lambda(x)$,

(c) π *is a Nash equilibrium of the game* $\mathcal{G}$,
(d) $\nabla_\pi \mathcal{L}(\pi,\mu)=0$ *for some* $\mu\in\mathbb{R}^N$,
(e) π *is a critical point of* Ψ *on* $\Delta(\mathcal{R})^N$, *i.e.* $\nabla\Psi(\pi)\perp\Delta_0^N$ *where* Δ_0 *is the tangent space to* $\Delta(\mathcal{R})$, *namely* $\Delta_0=\{z\in\mathbb{R}^M\colon \sum_{r\in\mathcal{R}} z^r=0\}$.

Proof. The equivalence (a) $\Leftrightarrow$ (b) is obvious if we note that (7) and (18) describe the same dynamics, while (a) $\Leftrightarrow$ (c) was proved in Proposition 3. The equivalence (d) $\Leftrightarrow$ (e) is also straightforward. For (a) $\Leftrightarrow$ (d) we observe that the vector μ in (d) is uniquely determined: indeed, the condition $\nabla_\pi\mathcal{L}(\pi,\mu)=0$ gives $\frac{\partial H}{\partial \pi^{ir}}(\pi)-\frac{1}{\beta_i}\ln(\pi^{ir})=\mu^i$ so that setting $x=\nabla H(\pi)$ we get $\pi^{ir}=\exp[\beta_i(x^{ir}-\mu^i)]$ and since $\pi^i\in\Delta(\mathcal{R})$ we deduce $\mu^i=\lambda^i(x^i)$. From this observation it follows that (d) may be equivalently expressed by the equations $x=\nabla H(\pi)$ and $\pi=\Sigma(x)$ which is precisely (8) and therefore (a) $\Leftrightarrow$ (d). □

When the quantities $\beta_i\delta$ are small, the function Ψ turns out to be concave so we may add another characterization of the equilibria and a weaker alternative condition for uniqueness.

Proposition 12. *Let* $\beta=\max_{i\in\mathcal{P}}\beta_i$. *If* $\beta\delta<1$ *then* Ψ *is strongly concave with parameter* $(\frac{1}{\beta}-\delta)$ *and attains its maximum at a unique point* $\bar\pi\in\Delta$. *This point* $\bar\pi$ *is the only Nash equilibrium of the game* $\mathcal{G}$ *while* $\bar x=F(\bar\pi)$ *is the corresponding unique rest point of the adaptive dynamics* (7).

Proof. It suffices to prove that $h'\nabla^2\Psi(\pi)h\leqslant-(\frac{1}{\beta}-\delta)\|h\|^2$ for all $h\in\Delta_0^N$. Using Lemma 9 we get

$$h'\nabla^2\Psi(\pi)h=\sum_{r\in\mathcal{R}}\Bigg[\sum_{i\neq j}h^{ir}h^{jr}\mathbb{E}_\pi^B\big[c^r_{U^r_{ij}+1}-c^r_{U^r_{ij}+2}\big]-\sum_i\frac{1}{\beta_i\pi^{ir}}\big(h^{ir}\big)^2\Bigg]. \tag{21}$$

Setting $Y^{ir}=v^{ir}X^{ir}$ with $v^{ir}=\frac{h^{ir}}{\pi^{ir}}$, and $\delta^r_u=(c^r_u-c^r_{u-1})$ with $\delta^r_0=\delta^r_1=0$, this may be rewritten as

$$\begin{aligned}
h'\nabla^2\Psi(\pi)h&=\sum_{r\in\mathcal{R}}\Bigg[\sum_{i\neq j}v^{ir}v^{jr}\pi^{ir}\pi^{jr}\mathbb{E}_\pi^B\big[c^r_{U^r_{ij}+1}-c^r_{U^r_{ij}+2}\big]-\sum_i\frac{\pi^{ir}}{\beta_i}\big(v^{ir}\big)^2\Bigg]\\
&=\sum_{r\in\mathcal{R}}\mathbb{E}_\pi^B\Bigg[\sum_{i\neq j}Y^{ir}Y^{jr}\big(c^r_{U^r-1}-c^r_{U^r}\big)-\sum_i\frac{1}{\beta_i}\big(Y^{ir}\big)^2\Bigg]\\
&\leqslant\sum_{r\in\mathcal{R}}\mathbb{E}_\pi^B\Bigg[-\delta^r_{U^r}\sum_{i\neq j}Y^{ir}Y^{jr}-\frac{1}{\beta}\sum_i\big(Y^{ir}\big)^2\Bigg]\\
&=\sum_{r\in\mathcal{R}}\mathbb{E}_\pi^B\Bigg[-\delta^r_{U^r}\Bigg(\sum_i Y^{ir}\Bigg)^2-\Bigg(\frac{1}{\beta}-\delta^r_{U^r}\Bigg)\sum_i\big(Y^{ir}\big)^2\Bigg].
\end{aligned}$$

The conclusion follows by neglecting the first term in the latter expectation and noting that $\delta^r_{U^r}\leqslant\delta$ while $\mathbb{E}[(Y^{ir})^2]=\frac{(h^{ir})^2}{\pi^{ir}}\geqslant(h^{ir})^2$. □

When the costs c^r_u are linear the previous result may be slightly improved.

Proposition 13. *Suppose that the route costs are linear* $c^r_u=a^r+\delta^r u$. *Let* $\beta=\max_{i\in\mathcal{P}}\beta_i$ *and* δ *given by* (14). *If* $\beta\delta<2$ *then the function* $\Psi(\cdot)$ *is quadratic and strongly concave on the space* $\Delta(\mathcal{R})^N$ *with parameter* $(\frac{2}{\beta}-\delta)$.

Proof. Under the linearity assumption Eq. (21) gives

$$\begin{aligned}
h'\nabla^2\Psi(\pi)h&=-\sum_{r\in\mathcal{R}}\Bigg[\delta^r\sum_{i\neq j}h^{ir}h^{jr}+\sum_i\frac{1}{\beta_i\pi^{ir}}\big(h^{ir}\big)^2\Bigg]\\
&=-\sum_{r\in\mathcal{R}}\Bigg[\delta^r\Bigg\{\Bigg(\sum_i h^{ir}\Bigg)^2-\sum_i\big(h^{ir}\big)^2\Bigg\}+\sum_i\frac{1}{\beta_i\pi^{ir}}\big(h^{ir}\big)^2\Bigg]\\
&\leqslant\sum_{r\in\mathcal{R}}\Bigg[\delta\sum_i\big(h^{ir}\big)^2-\frac{1}{\beta}\sum_i\frac{1}{\pi^{ir}}\big(h^{ir}\big)^2\Bigg].
\end{aligned}$$

Maximizing this latter expression with respect to the variables $\pi^{ir}\geqslant0$ under the constraints $\sum_r\pi^{ir}=1$, we get

$$h'\nabla^2\Psi(\pi)h \leqslant \delta\sum_i\sum_r (h^{ir})^2 - \frac{1}{\beta}\sum_i\Big(\sum_r |h^{ir}|\Big)^2 = \sum_i\Big[\delta\|h^i\|_2^2 - \frac{1}{\beta}\|h^i\|_1^2\Big].$$

Now if we restrict to vectors $h=(h^i)_{i\in\mathcal{P}}$ in the tangent space Δ_0^N, that is to say, $\sum_r h^{ir}=0$, we may use the inequality $\|h^i\|_1 \geqslant \sqrt{2}\|h^i\|_2$ to conclude

$$h'\nabla^2\Psi(\pi)h \leqslant \sum_i\Big[\delta\|h^i\|_2^2 - \frac{2}{\beta}\|h^i\|_2^2\Big] = -\Big(\frac{2}{\beta}-\delta\Big)\|h\|_2^2. \qquad \square$$

The characterization of $\bar\pi$ as a minimizer suggests that Ψ might provide an alternative Lyapunov function to study the asymptotic convergence under weaker assumptions than Theorem 10. Unfortunately, numerical simulations show that neither the energy $\Psi(\pi)$ nor the potential $H(\pi)$ decrease along the trajectories of (7), at least initially. However they do decrease for large t and therefore they may eventually serve as local Lyapunov functions near $\bar\pi$.

3.3. The symmetric case

In this final section we consider the case in which all players are identical with $\beta_i\equiv\beta$ for all $i\in\mathcal{P}$. We denote $\sigma(\cdot)$ the common Logit function (9). Under these circumstances one might expect rest points to be also symmetric with all players sharing the same perceptions: $\bar x^i=\bar x^j$ for all $i,j\in\mathcal{P}$. This is indeed the case when $\beta\delta$ is small, but beyond a certain threshold there is a multiplicity of rest points all of which except for one are non-symmetric.

Lemma 14. *For all $x,y\in\Omega$, each $i,j\in\mathcal{P}$ and every $r\in\mathcal{R}$, we have*

$$\big|C^{ir}(x)-C^{jr}(x)\big| \leqslant \frac{1}{2}\beta\delta\big\|x^i-x^j\big\|_\infty. \tag{22}$$

Proof. We observe that the only difference between F^{ir} and F^{jr} is an exchange of π^{ir} and π^{jr}. Thus, Proposition 8 and Lemma 9 combined imply that $|F^{ir}(\pi)-F^{jr}(\pi)|\leqslant\delta|\pi^{ir}-\pi^{jr}|$ and then (22) follows from the equality $C(x)=F(\Sigma(x))$ and Lemma 6. $\square$

Theorem 15. *If $\beta_i\equiv\beta$ for all $i\in\mathcal{P}$ then the adaptive dynamics (7) has exactly one symmetric rest point $\hat x=(\hat y,\dots,\hat y)$. Moreover, if $\beta\delta<2$ then every rest point is symmetric (thus unique).*

Proof. *Existence*. Consider the continuous map T from the cube $\prod_{r\in\mathcal{R}}[-c_N^r,-c_1^r]$ to itself that maps y to $T(y)=(T^r(y))_{r\in\mathcal{R}}$ where $T^r(y)=C^{ir}(y,\dots,y)$. Brouwer's theorem implies the existence of a fixed point $\hat y$ so that setting $\hat x=(\hat y,\dots,\hat y)$ we get a symmetric rest point for (7).

Uniqueness. Suppose $\hat x=(\hat y,\dots,\hat y)$ and $\tilde x=(\tilde y,\dots,\tilde y)$ are two distinct symmetric rest points and assume with no loss of generality that the set $\mathcal{R}^+=\{r\in\mathcal{R}:\ \tilde y^r<\hat y^r\}$ is non-empty. Let $\mathcal{R}^-=\mathcal{R}\setminus\mathcal{R}^+$. The fixed point condition gives $C^{ir}(\tilde x)<C^{ir}(\hat x)$ for all $r\in\mathcal{R}^+$, and since F^{ir} is decreasing with respect to the probabilities π^{jr} we deduce $\sigma^r(\tilde y)>\sigma^r(\hat y)$. Summing over all $r\in\mathcal{R}^+$ and setting $Q(z)=[\sum_{a\in\mathcal{R}^-}e^{\beta z^a}]/[\sum_{a\in\mathcal{R}^+}e^{\beta z^a}]$ we get

$$\frac{1}{1+Q(\tilde y)} = \sum_{r\in\mathcal{R}^+}\sigma^r(\tilde y) > \sum_{r\in\mathcal{R}^+}\sigma^r(\hat y) = \frac{1}{1+Q(\hat y)}$$

and therefore $Q(\hat y)>Q(\tilde y)$. However, $e^{\beta\hat y^r}>e^{\beta\tilde y^r}$ for $r\in\mathcal{R}^+$ and $e^{\beta\hat y^r}\leqslant e^{\beta\tilde y^r}$ for $r\in\mathcal{R}^-$, so that $Q(\hat y)<Q(\tilde y)$ which yields a contradiction.

Symmetry. Suppose next that $\beta\delta<2$ and let x be any rest point. For any two players $i,j\in\mathcal{P}$ and all routes $r\in\mathcal{R}$, property (22) gives

$$\big|x^{ir}-x^{jr}\big| = \big|C^{ir}(x)-C^{jr}(x)\big| \leqslant \frac{1}{2}\beta\delta\big\|x^i-x^j\big\|_\infty$$

and then $\|x^i-x^j\|_\infty\leqslant\frac{1}{2}\beta\delta\|x^i-x^j\|_\infty$ which implies $x^i=x^j$. $\square$

Corollary 16. *If $\beta_i\equiv\beta$ for all $i\in\mathcal{P}$ then the game $\mathcal{G}$ has a unique symmetric equilibrium. Moreover, if $\beta\delta<2$ then every equilibrium is symmetric (hence unique).*

The existence of a symmetric rest point requires not only that players be identical in terms of the β_i's but also with respect to payoffs. If these payoffs are given by $C^{ir}(x)=C^r(x)+\alpha^{ir}$ where $C^r(x)$ is a common value which depends only on the number of players that use route r and α^{ir} is a user specific value, then symmetry may be lost.

Going back to the stability of rest points we observe that the condition $\omega\delta < 2$ in Theorem 10 becomes more and more stringent as the number of players grows: for identical players this condition reads $\beta\delta < \frac{2}{N-1}$. Now, since $\beta\delta < 2$ already guarantees a unique rest point $\hat{x}$, one may expect that this remains an attractor under this weaker condition. Although extensive numerical experiments confirm this conjecture, we have only been able to prove that $\hat{x}$ is a *local* attractor. Unfortunately this does not allow to conclude the almost sure convergence of the learning process (2).

Theorem 17. *If $\beta_i \equiv \beta$ for all $i \in \mathcal{P}$ with $\beta\delta < 2$ then the unique rest point $\hat{x} = (\hat{y}, \dots, \hat{y})$ is symmetric and a local attractor for the dynamics* (7)*.*

Proof. We will prove that

$$\Phi(x) = \max\left\{\max_{i,j}\|x^i - x^j\|_\infty, \frac{1}{N-1}\max_i\|x^i - \hat{y}\|\right\}$$

is a local Lyapunov function. More precisely, fix any $\varepsilon > 0$ and choose a lower bound $\bar{\pi} \leqslant \pi^{ir} := \sigma^r(x^i)$ over the compact set $S_\varepsilon = \{\Phi \leqslant \varepsilon\}$. This set is a neighborhood of $\hat{x}$ since the minimum of Φ is attained at $\Phi(\hat{x}) = 0$. Now set $\alpha = \frac{1}{2}\beta\delta$ and $b = \bar{\pi}[1-\alpha]$, and reduce ε so that $\beta|C^{jr}(x) - x^{jr}| \leqslant b$ for all $x \in S_\varepsilon$. We claim that S_ε is invariant for the dynamics with $\rho(t) = \Phi(x(t))$ decreasing to 0. To this end we show that $\dot{\rho}(t) \leqslant -\frac{b}{2}\rho(t)$. We compute $\dot{\rho}(t)$ distinguishing 3 cases.

Case 1: $x^{ir} - x^{jr} = \rho(t)$.
A simple manipulation using (7) gives

$$\dot{x}^{ir} - \dot{x}^{jr} = -\pi^{ir}\rho(t) + \pi^{ir}\big[C^{ir}(x) - C^{jr}(x)\big] + \big(\pi^{ir} - \pi^{jr}\big)\big[C^{jr}(x) - x^{jr}\big]$$

so that (22) implies

$$\dot{x}^{ir} - \dot{x}^{jr} \leqslant -\pi^{ir}[1-\alpha]\rho(t) + \frac{1}{2}\beta\big|C^{jr}(x) - x^{jr}\big|\rho(t) \leqslant -\frac{b}{2}\rho(t).$$

Case 2: $\frac{1}{N-1}(x^{ir} - \hat{y}^r) = \rho(t)$.
In this case we have $x^{ia} - \hat{y}^a \leqslant x^{ir} - \hat{y}^r$ for all $a \in \mathcal{R}$ so that

$$\sigma^r\big(x^i\big) = \left[\sum_{a\in\mathcal{R}} e^{\beta(x^{ia} - x^{ir})}\right]^{-1} \geqslant \left[\sum_{a\in\mathcal{R}} e^{\beta(\hat{y}^a - \hat{y}^r)}\right]^{-1} = \sigma^r(\hat{y})$$

which then implies

$$C^{ir}\big(x^i, \dots, x^i\big) \leqslant C^{ir}(\hat{y}, \dots, \hat{y}) = \hat{y}^r = x^{ir} - (N-1)\rho(t).$$

On the other hand, using Lemmas 9 and 6 we have

$$\big|C^{ir}(x) - C^{ir}\big(x^i, \dots, x^i\big)\big| \leqslant (N-1)\alpha\max_{j\neq i}\|x^j - x^i\|_\infty \leqslant (N-1)\alpha\rho(t) \tag{23}$$

so that $C^{ir}(x) - x^{ir} \leqslant -(N-1)[1-\alpha]\rho(t)$ and therefore

$$\frac{d}{dt}\left[\frac{1}{N-1}\big(x^{ir} - \hat{y}^r\big)\right] \leqslant -\pi^{ir}[1-\alpha]\rho(t) \leqslant -\frac{b}{2}\rho(t).$$

Case 3: $\frac{1}{N-1}(\hat{y}^r - x^{ir}) = \rho(t)$.
Similarly to the previous case we now have $\sigma^r(x^i) \leqslant \sigma^r(\hat{y})$ so that

$$C^{ir}\big(x^i, \dots, x^i\big) \geqslant C^{ir}(\hat{y}, \dots, \hat{y}) = \hat{y}^r = x^{ir} + (N-1)\rho(t).$$

This combined with (23) gives $x^{ir} - C^{ir}(x) \leqslant -(N-1)[1-\alpha]\rho(t)$ and then as in the previous case we deduce

$$\frac{d}{dt}\left[\frac{1}{N-1}\big(\hat{y}^r - x^{ir}\big)\right] \leqslant -\frac{b}{2}\rho(t). \qquad \square$$

This last result shows that $\hat{x}$ is a local attractor when $\beta\delta < 2$. As a matter of fact, in this case we have observed through extensive simulations that (7) always converges towards $\hat{x}$ so we conjecture that it is a global attractor. More generally, the simulations show that even for values $\beta\delta > 2$ the continuous dynamics converge towards an equilibrium, although there is a bifurcation value beyond which the symmetric equilibrium becomes unstable and convergence occurs towards one of the multiple non-symmetric equilibria. The structure of the bifurcation is quite intricate and deserves more attention. The possibility of selecting the equilibrium attained by controlling the payoffs using tolls or delays to incorporate the externality that each user imposes to the rest, may be also of interest in this context. Eventually one might think of a feedback mechanism in the adaptive dynamics that would lead the system to a more desirable equilibrium from the point of view of the planner.

Acknowledgment

We warmly thank Jaime San Martín for stimulating discussions on *sums of Bernoullis* which led to Proposition 12. Sylvain Sorin was supported by ANR-05-BLAN-0248-01. He acknowledges apt comments from Josef Hofbauer and Ed Hopkins during the conference "Evolutionary Game Dynamics-BIRS" at Banff in June 2006. We also thank two anonymous referees and the Associate Editor for constructive suggestions that helped to improve the final presentation of the paper.

References

Arthur, W., 1993. On designing economic agents that behave like human agents. J. Evolutionary Econ. 3, 1–22.
Auer, P., Cesa-Bianchi, N., Freud, Y., Schapire, R., 2002. The non-stochastic multiarmed bandit problem. SIAM J. Comput. 32 (1), 48–77.
Avineri, E., Prashker, J.N., 2006. The impact of travel time information on travellers' learning under uncertainty. Transportation 33, 393–408.
Beggs, A., 2005. On the convergence of reinforcement learning. J. Econ. Theory 122, 1–36.
Benaim, M., 1999. Dynamics of stochastic approximation algorithms. In: Séminaire de Probabilités. In: Lecture Notes in Math., vol. 1709. Springer, Berlin, pp. 1–68.
Benaim, M., Hirsch, M., 1999. Mixed equilibria and dynamical systems arising from fictitious play in perturbed games. Games Econ. Behav. 29, 36–72.
Benaim, M., Hofbauer, J., Sorin, S., 2005. Stochastic approximations and differential inclusions. SIAM J. Control Optim. 44, 328–348.
Borgers, T., Sarin, R., 1997. Learning through reinforcement and replicator dynamics. J. Econ. Theory 77, 1–14.
Brown, G., 1951. Iterative solution of games by fictitious play. In: Activity Analysis of Production and Allocation. In: Cowles Commission Monograph, vol. 13. John Wiley & Sons, Inc., New York, pp. 374–376.
Cantarella, G., Cascetta, E., 1995. Dynamic processes and equilibrium in transportation networks: Towards a unifying theory. Transp. Sci. 29, 305–329.
Cascetta, E., 1989. A stochastic process approach to the analysis of temporal dynamics in transportation networks. Transp. Res. 23B, 1–17.
Daganzo, C., Sheffi, Y., 1977. On stochastic models of traffic assignment. Transp. Sci. 11, 253–274.
Davis, G., Nihan, N., 1993. Large population approximations of a general stochastic traffic model. Oper. Res. 41 (1), 170–178.
Dial, R., 1971. A probabilistic multipath traffic assignment model which obviates path enumeration. Transp. Res. 5, 83–111.
Duffy, J., Hopkins, E., 2005. Learning, information, and sorting in market entry games: Theory and evidence. Games Econ. Behav. 51, 31–62.
Erev, I., Roth, A., 1998. Predicting how people play games: Reinforcement learning in experimental games with a unique, mixed strategy equilibria. Amer. Econ. Rev. 88, 848–881.
Foster, D., Vohra, R., 1997. Calibrated learning and correlated equilibria. Games Econ. Behav. 21, 40–55.
Foster, D., Vohra, R., 1998. Asymptotic calibration. Biometrika 85, 379–390.
Freund, Y., Schapire, R., 1999. Adaptive game playing using multiplicative weights. Games Econ. Behav. 29 (1), 79–103.
Friesz, T., Berstein, D., Mehta, N., Tobin, R., Ganjalizadeh, S., 1994. Day-to-day dynamic network disequilibria and idealized traveler information systems. Oper. Res. 42, 1120–1136.
Fudenberg, D., Levine, D., 1998. The Theory of Learning in Games. Series on Economic Learning and Social Evolution, vol. 2. MIT Press, Cambridge, MA.
Hannan, J., 1957. Approximation to Bayes risk in repeated plays. In: Dresher, M., Tucker, A.W., Wolfe, P. (Eds.), Contributions to the Theory of Games. Princeton Univ. Press, pp. 97–139.
Hart, S., 2005. Adaptive heuristics. Econometrica 73, 1401–1430.
Hart, S., Mas-Collel, A., 2001. A reinforcement procedure leading to correlated equilibrium. In: Debreu, G., Neuefeind, W., Trockel, W. (Eds.), Economics Essays: A Festschrift for Werner Hildebrand. Springer, Berlin.
Hofbauer, J., Sandholm, W., 2002. On the global convergence of stochastic fictitious. Econometrica 70, 2265–2294.
Horowitz, J., 1984. The stability of stochastic equilibrium in a two-link transportation network. Transp. Res. Part B 18, 13–28.
Kushner, H., Yin, G., 1997. Stochastic Approximations Algorithms and Applications. Appl. Math., vol. 35. Springer-Verlag, New York.
Laslier, J.-F., Topol, R., Walliser, B., 2001. A behavioral learning process in games. Games Econ. Behav. 37, 340–366.
McKelvey, R., Palfrey, T., 1995. Quantal response equilibria for normal form games. Games Econ. Behav. 10, 6–38.
Monderer, D., Shapley, L., 1996. Potential games. Games Econ. Behav. 14, 124–143.
Posch, M., 1997. Cycling in a stochastic learning algorithm for normal form games. J. Evol. Econ. 7, 193–207.
Robinson, J., 1951. An iterative method of solving a game. Ann. of Math. 54, 296–301.
Rosenthal, R., 1973. A class of games possessing pure-strategy Nash equilibria. Internat. J. Game Theory 2, 65–67.
Sandholm, W., 2001. Potential games with continuous player sets. J. Econ. Theory 97, 81–108.
Sandholm, W., 2002. Evolutionary implementation and congestion pricing. Rev. Econ. Stud. 69, 667–689.
Selten, R., Guth, W., 1982. Equilibrium point selection in a class of market entry games. In: Diestler, M., Furst, E., Schwadiauer, G. (Eds.), Games Economics Dynamics and Time Series Analysis. Physica-Verlag, Wien–Wurzburg.
Selten, R., Chmura, T., Pitz, T., Kube, S., Schreckenberg, M., 2007. Commuters route choice behaviour. Games Econ. Behav. 58, 394–406.
Smith, M., 1984. The stability of a dynamic model of traffic assignment: An application of a method of Lyapunov. Transp. Sci. 18, 245–252.
Wardrop, J., 1952. Some theoretical aspects of road traffic research. Proc. Inst. Civil Eng. 1, 325–378.
Young, P., 2004. Strategic Learning and Its Limits. Oxford University Press.