

# Learning and equilibrium selection in a coordination game with heterogeneous agents

Alberto Fogale<sup>a</sup>, Paolo Pellizzari<sup>b,\*</sup>, Massimo Warglien<sup>c</sup>

<sup>a</sup>*Department of Applied Mathematics (DMA), “Ca’ Foscari” University, Dorsoduro 3825/E, 30123 Venice, Italy*

<sup>b</sup>*DMA and School of Advanced Studies in Venice (SSAV), Dorsoduro 3488, 30123 Venice, Italy*

<sup>c</sup>*SSE and Department of Economics and Business Management, “Ca’ Foscari” University, Cannaregio 873, 30121 Venice, Italy*

Received 8 May 2006; received in revised form 31 December 2006

Available online 1 March 2007

---

## Abstract

We study a modified version of the coordination game presented in [J. van Huyck, J. Cook, R. Battalio, Selection dynamics, asymptotic stability, and adaptive behavior, *J. Polit. Econ.* 102 (1994) 975–1005], where a representative selection dynamics was proposed to explain experimental data. Assuming that the agents adjust their moves in the direction of the best response, we derive a formal analysis of the stability of the equilibria. We show by simulation that the interior equilibrium is robustly reached even when considerable heterogeneity is allowed among the agents. Our truly multi-agent game is capable of approximating quite well both the “median” game convergence and the experimental data. © 2007 Elsevier B.V. All rights reserved.

*Keywords:* Coordination games; Selection dynamics; Equilibrium selection; Heterogeneous agents

---

## 1. Introduction

Whenever multiple equilibria arise in a game, the usefulness of the model is considerably hampered unless some criterion is found to select among competing solutions. Learning—or broadly speaking adaptive behavior—has often been advocated as a candidate mechanism for equilibrium selection in games. It has been argued that a plausible criterion for equilibrium selection is that an equilibrium should be learnable by agents employing some adaptive rule of behavior and be (locally) stable under the interaction dynamics generated by such rule [1,2]. Tighter requirements for equilibrium selection are emerging as a large laboratory evidence on learning in experimental games is being cumulated [3]. Such evidence suggests to introduce at least two additional requirements for criteria of equilibrium selection.

1. Beyond asymptotic stability, the convergence speed of the adaptive rules to equilibrium should be compatible with empirical convergence times.

---

\*Corresponding author. Tel.: +39 41 2346924; fax: +39 41 5221756.

E-mail addresses: [albertofogale@gmail.com](mailto:albertofogale@gmail.com) (A. Fogale), [paolop@unive.it](mailto:paolop@unive.it) (P. Pellizzari), [warglien@unive.it](mailto:warglien@unive.it) (M. Warglien).

2. Dynamic stability of the adaptive rules should be robust enough to tolerate some heterogeneity in agents learning behavior [3]

In this paper, we address these issues in the context of a coordination game belonging to a class of games formerly explored by van Huyck et al. [4]. Problems of equilibrium selection arise naturally in coordination games, since the multiplicity of equilibria is one of their defining features [5]. van Huyck et al. [4] have studied both experimentally and analytically a simple coordination game called  $\Gamma(w)$ .

The game has two equilibria, a corner one and an interior one. Careful laboratory experiments conducted by van Huyck et al. show that human subjects converge in fairly short time to the interior equilibrium for both low and high values of  $w$ , while they never converge to the corner equilibrium. A simple and classical adaptive rule, the Cournot myopic best-response dynamics, matches these results for values of  $w$  ranging in the (1,3] interval, since it implies the stability of the interior equilibrium but the instability of the corner one. On the other hand, for  $w > 3$ , both equilibria are unstable under myopic Cournot dynamics, in sharp contrast with the experimental evidence. However, van Huyck et al. show that a rule of partial, inertial adaptation with adjustment strength decaying over time (referred to as the L-map, from Lucas) predicts that even for values of  $w$  in the (3, 4] interval, the corner equilibrium is unstable while the interior one is stable, thus reconciling model and data. They conclude that the L-map provides an accurate selection criterion for the  $\Gamma(w)$  game.

While van Huyck et al.'s paper is a remarkable effort of combining theoretical insights and experimental results, it suffers from the important limitation that their formal analysis is conducted hypothesizing a “representative agent” that, in fact, has to learn to coordinate with himself (the adaptive rule directly “decides” the median behavior in time). Well-known problems with the use of the representative agent are especially acute when coordination problems are implied [6]. Chen et al. [7] try to overcome this limitation by studying adaptation in the  $\Gamma(w)$  game through the genetic programming paradigm, which allows to simulate multi-agent learning with a heterogeneity of individual behaviors. While their simulations show that genetic programming selects the interior equilibrium for both high and low  $w$  values, convergence time is much slower than the experimentally observed one, and there are persisting fluctuations around the equilibrium that are not found in experimental data. More fundamentally, while genetic programming provides a powerful machine learning algorithm, its psychological plausibility as a model of human learning is questionable.

We undertake a different way by considering a slight variant of the  $\Gamma(w)$  game, which allows a formal analysis of learning dynamics without resorting to the representative agent fiction: we substitute the median action  $M(e)$  with the average action  $A(e)$ . Furthermore, we consider a learning rule which is a generalization of both myopic best-response and the L-map, and is well grounded in the psychological literature.

The paper is organized as follows. In Section 2 we briefly review the coordination game of van Huyck et al., discuss the use of simple average instead of the median in the payoff function and provide some justification for the use of our adaptive learning rule. Section 3 discusses the similarities of our learning rule with Cournot myopic best-response and L-map adaptation. Moreover, we prove analytically some stability properties of the modified game. In a nutshell, the interior equilibrium is robustly selected for a wide interval of adjustment parameters. Numerical experiments and a comparison with original experimental data of van Huyck et al. are presented in Section 4. Finally, some conclusive remarks are given in the last section.

## 2. The coordination game

We consider the coordination problem described in Ref. [4] in which  $n$  agents play a game  $\Gamma(w)$  whose payoffs  $\pi$  is depending on the median action  $M(e)$ :

$$\pi(e^i, e^{-i}) = c_1 - c_2 |e^i - wM(e)(1 - M(e))|,$$

where  $w \in (1, 4]$ ,  $c_1, c_2 > 0$  are parameters,  $e^i \in [0, 1]$  denotes the action of player  $i$ ,  $e^{-i} = (e^1, \dots, e^{i-1}, e^{i+1}, \dots, e^n)$ . The payoff is clearly maximal when  $e^i = wM(e)(1 - M(e))$ , i.e., when  $i$ th agent plays an action that is a function of the median action of the others. It is straightforward to notice that there are two symmetric Nash equilibria, namely those in which all players chose, respectively, actions 0 or  $1 - 1/w$  that are fixpoints of the best-response function  $b(M) = wM(1 - M)$ . van Huyck et al. focus on the cases  $w = 2.47222$  and 3.86957. These two values will be named “low” and “high” in the sequel. While the corner equilibrium in

which all players chose 0 is unstable for both values of  $w$ , the interior equilibrium obtained when all players choose  $1 - 1/w$  turns out to be stable for  $w = 2.47222$  and unstable when  $w = 3.86957$ . As any equilibrium can be interpreted as a potential convention among agents, it is of interest to study selection dynamics that converge to the equilibrium. van Huyck et al. consider a family of “representative” selection dynamics described by the dynamical system

$$M_t = b(M_t^e), \quad (1)$$

$$M_t^e = M_{t-1}^e + \alpha_{t-1}(M_{t-1} - M_{t-1}^e) \quad (2)$$

for  $t \geq 2$ ,  $0 < \alpha_t \leq 1$ ;  $M_t^e$  is the expectation of the representative agent. It is shown that the myopic best-response obtained setting  $\alpha_t = 1$  for all  $t$  converges to the interior equilibrium only for low  $w$ . On the other hand, some inertia (or adaptation) is beneficial to stabilize the interior equilibrium and the position  $\alpha_t = 1/t$  allows to get convergence even when  $w$  is high. This is in perfect agreement with the experimental evidence provided in the paper, where a number of experiments with real subjects are described. In all cases (for different values of  $w$ ) humans coordinate on the interior equilibrium. The authors then conclude that “The inertial selection dynamics accurately predicts the behavior observed in our experiment. The myopic best-response does not”.

In order to avoid the pitfalls of a “representative agent” model, that is implausible from both psychological and economic standpoints, we undertake a different way by considering the game  $\Gamma'(w)$ , a slight variant of  $\Gamma(w)$ , which allows a formal analysis of learning dynamics without resorting to the representative agent fiction.

Thus, we consider the same game as in van Huyck, substituting the median action  $M(e)$  with the average action  $A(e)$  for analytical convenience, setting:

$$\pi'(e_i, e_{-i}) = c_1 - c_2|e_i - wA(e)(1 - A(e))|.$$

The substitution we propose needs some further justification. On the one hand, the use of median values rather than average ones in experimental coordination games can be explained by the fact that, given the small number of participants to each experimental session, the median action is more robust to the presence of outliers than the average action. However, in an analytic context such robustness considerations lose much of their relevance. On the other hand, we show in Section 3 that in the average action game the adaptive rules considered by van Huyck et al. preserve (even in a multi-agent setting) the same stability properties demonstrated in the median action game.

In an authentic multi-agent framework, instead of a representative agent learning rule, we consider individual learning rules and study their simultaneous dynamics. The simple individual rule for the  $i$ th agent we analyze is

$$e_{i,t+1} = (1 - \lambda_i)e_{i,t} + \lambda_i b(A_t(e)), \quad (3)$$

where  $b(A_t(e))$  is the best response to the average action at time  $t$  and  $\lambda_i > 0$  is a parameter determining the individual speed of learning. The rule turns out to be an instantiation of the more general “learning direction theory” of Selten and Stocker [8] and Selten et al. [9]. In the learning direction theory (LDT), agents adjust in the direction of the ex post best response to last period moves of other players. While the LDT pays the price of having no parameters by providing only qualitative predictions, our rule introduces one parameter (the speed of learning) but allows more precise predictions.<sup>1</sup> A simple test conducted on van Huyck et al. experimental data along the line described in Selten’s work shows a strong agreement with the LDT, providing indirect support to our learning rule. Table 1 exhibits the results of a test of the null hypothesis of random behavior against the alternative that the players follow some form of LDT. The null is rejected at conventional significance level in seven instances out of eight, in favor of a behavior compatible with the LDT. As the learning rule (3) always agrees with LDT, we interpret these findings as supporting<sup>2</sup> the use of (3).

<sup>1</sup>The rule also bears close resemblance with the Rescorla–Wagner model of learning, a model supported by a broad empirical evidence in experimental psychology, see Refs. [10,11].

<sup>2</sup>This evidence is somehow indirect as other rules might be compatible with such results.

Table 1

The table shows the  $p$ -values of Wilcoxon test run on the experimental data of van Huyck et al. [4] for  $w = 3.86957$

Wilcoxon test	
Session	$p$ -value
1	0.039
2	0.033
3	0.021
4	0.021
5	0.021
6	0.039
7	0.021
8	0.072

The null hypothesis of random behavior of agents can be rejected in favor of LDT in seven cases out of eight at the confidence level of 5% despite rather small sample sizes.

### 3. A formal analysis

This section presents a formal treatment of the average game along the following steps. We first study the stability of the fixpoints of  $\Gamma'(w)$  assuming that all agents have the same speed of learning  $\lambda$ . We then use a continuity argument to state that the same stability properties hold when agents are heterogeneous provided that their  $\lambda$ 's are not too different. It is of obvious interest to understand how much heterogeneity is allowed before stability breaks down. The answer is deferred to the following section where we show with numerical experiments that (3) is stable over a wide mixture of the  $\lambda$  individual parameters.

The following theorem gives conditions for the stability of the equilibria of the  $\Gamma'(w)$  game when all agents have the same speed of learning.

**Theorem 1.** *Assume  $\lambda_i \equiv \lambda > 0$  for all  $i$ . Then the average game admits, for any  $\lambda$ , an unstable corner equilibrium  $(0, \dots, 0)$ . The interior equilibrium*

$$\bar{e} = \left( 1 - \frac{1}{w}, 1 - \frac{1}{w}, \dots, 1 - \frac{1}{w} \right)$$

*is stable in the interval  $0 < \lambda < 2/(w - 1)$ .*

**Proof.** See Appendix.

The previous theorem shows that there is a wide set of individual  $\lambda$ 's that allows convergence to the interior equilibrium. In particular, considering the  $w$  values analyzed in van Huyck et al., if  $w = 2.47222$  then  $\bar{e}$  is stable for  $0 < \lambda < 1.353$ ; if  $w = 3.86957$  then  $\bar{e}$  is stable for  $0 < \lambda < 0.697$ . For low  $w$ , stability of the interior equilibrium holds even under overreaction<sup>3</sup> by agents ( $\lambda > 1$ ). When  $w = 3.86957$  the interior equilibrium is stable provided that agents do not adjust too much their previous action in the direction of the observed best-reply. The following result proves that the stability properties still hold in the presence of heterogeneous agents.

**Lemma 1.** *Consider the average game  $\Gamma'(w)$  and let  $\bar{\lambda}$  be such that  $0 < \bar{\lambda} < 2/(w - 1)$ . Then, if all agents have learning speed  $\lambda$  in a proper neighborhood of  $\bar{\Lambda} = (\bar{\lambda}, \bar{\lambda}, \dots, \bar{\lambda})$ , the interior equilibrium of the game is stable.*

**Proof.** The eigenvalues of the Jacobian matrix at the equilibrium are continuous functions of the learning rates  $\Lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$  of the heterogeneous agents. Hence there exists  $\Lambda$  in a neighborhood of  $\bar{\Lambda}$  such that all the components of the interior equilibrium are in  $[0, 1]$  and all the moduli of the eigenvalues of the Jacobian matrix are bounded away from 1. It follows immediately that the equilibrium is stable.  $\square$

<sup>3</sup>If  $\lambda > 1$ , an agent whose action, say, is  $e_t < b(A_t(e))$  will play  $e_{t+1} > A_t(e)$  in the following step, moving in the right direction even beyond the best previous response. In this sense, it is overreacting.

The theorem intuitively states that the selection dynamics of individual learning rule (3) is indeed capable of driving moderately heterogeneous agents to the interior equilibrium. Notice that this stability result holds even when variable  $\lambda$ 's are used by agents, provided that they eventually fall below an appropriate constant, 0.697 in our case. This is always the case when the  $\lambda$ 's have, say, the form  $1/t$ .

The result is novel in that there is no need for a representative agent and it sheds light on the individual process used by the agents in their effort to coordinate. Further exploration of this result can be pursued by resorting to numerical experiments, see for a similar example Ref. [12].

#### 4. Applications and numerical experiments

This section builds on the previous results to show by computer experiments the degree of heterogeneity that is allowed in the set of agents'  $\lambda$ 's. We also compare the multi-agent average and median games, showing numerically that the dynamics of the former closely approximates the latter (that defies an analytical treatment). All our experiments are conducted setting  $w = 3.86957$  in order to make them comparable with the original laboratory results of van Huyck et al. We remind that, in order to achieve stability, the corresponding  $\lambda$ 's upper bound is 0.697.

We run two kinds of computer experiments. In the first case, we sample the  $\lambda$ 's of the agents in the intervals  $(0, l)$  for 10 values of  $l \in \{0.11, 0.22, \dots, 0.88, 0.99\} \cup \{0.697\}$ . We simulate a population of 5 agents as in Ref. [4]. Clearly, values of  $l > 0.697$  do not allow the application of Lemma 1 to ensure convergence. In the second series of experiments, the  $\lambda$ 's of  $k$  agents out of 5 are sampled in  $(0, 0.697)$  while the remaining  $5 - k$  have coefficients in  $(0.697, 1)$ . This is intended to show the degree of "unstable" heterogeneity that is compatible with convergence. In order to get meaningful results, we run 100 trials of 20 periods for each experiment, displaying the mean result. To measure the distance from the equilibrium, we compute the distance  $|e^* - \frac{1}{5}\sum_i e_i|$  for each  $t$ , where  $e^* = 1 - 1/w$ . In Fig. 1 the distance from the equilibrium when all  $\lambda$ 's are sampled in  $[0, l]$  is displayed. The picture shows that, provided that  $\lambda_i \leq l = 0.66$ , convergence is quite fast, despite the heterogeneity of agents. Notice that when  $l$  increases some cycles are appearing, in analogy with the loss of stability of the logistic map.

The same conclusion can be inferred by looking at Fig. 2, depicting the distribution of the mean distance from equilibrium for  $10 \leq t \leq 20$ , in various cases. Most values are close to zero for values of  $l \leq 0.66$ , showing that less than 10 iterations suffice to reach equilibrium. The convergence is slower when  $l = 0.697$  and is never reached for larger values.

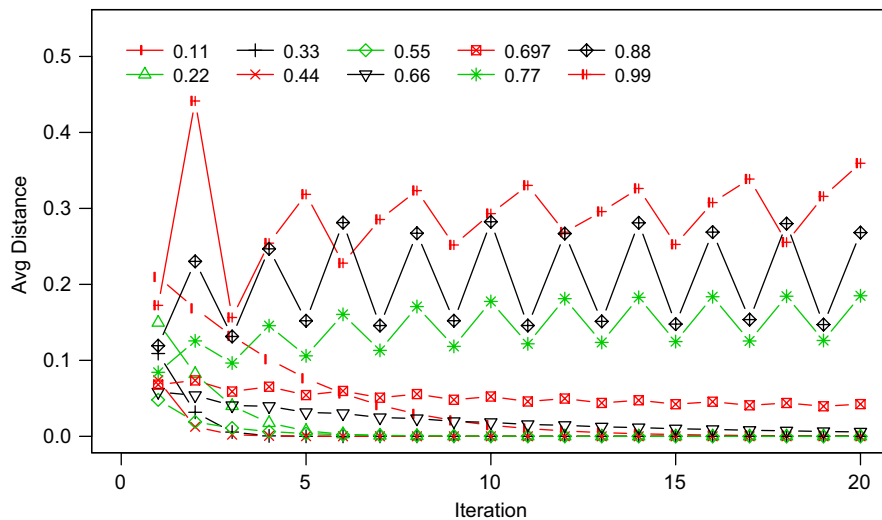


Fig. 1. Dynamics of average distance from equilibrium for various values of  $l$ , when  $\lambda$  is randomly sampled in  $[0, l]$ . The mean of 100 trials is depicted.

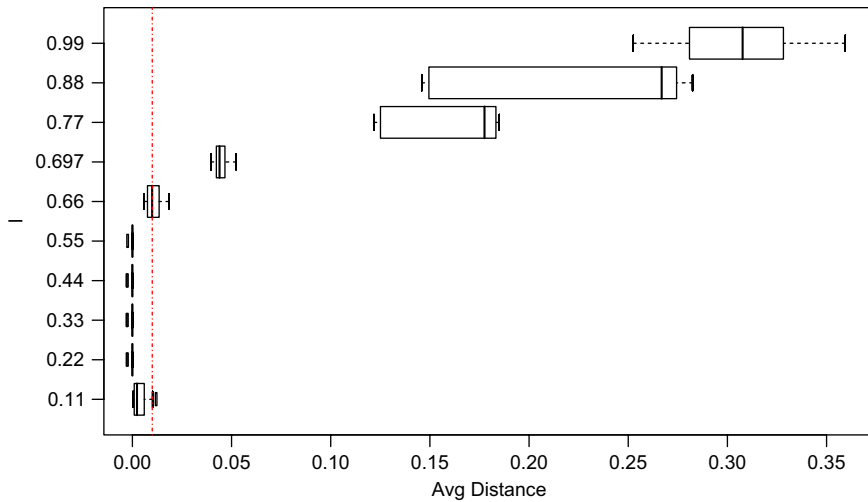


Fig. 2. Boxplots of distribution of average distances from the interior equilibrium for  $10 \leq t \leq 20$ , and different values of  $l$ .

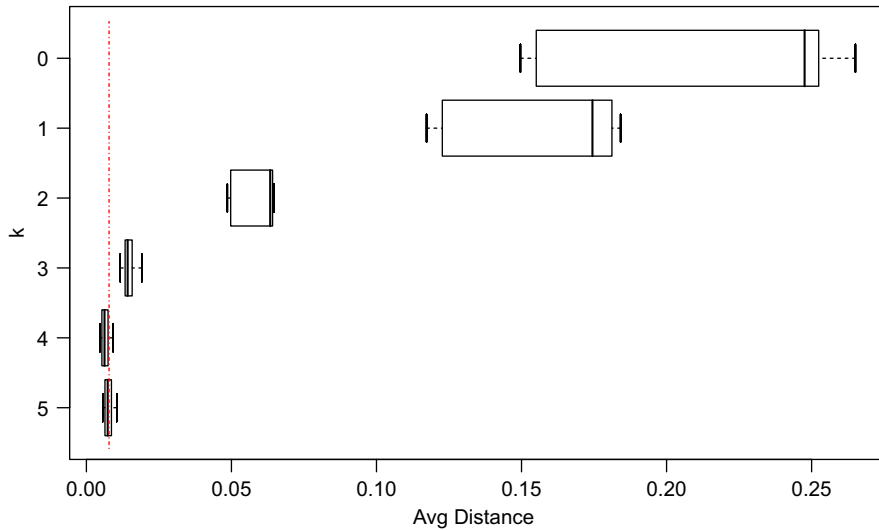


Fig. 3. Boxplots of distribution of average distances from the interior equilibrium for  $10 \leq t \leq 20$ , when  $k$  agents have  $\lambda \in (0, 0.697)$ .

We might probe Lemma 1 sampling  $\lambda$  in different stable and unstable intervals. Fig. 3 shows the distribution of the average distances from equilibrium for  $10 \leq t \leq 20$  when  $k$  out of 5 agents have  $\lambda$  sampled in  $(0, 0.697)$ . This is a situation where  $5 - k$  players overreact selecting  $\lambda$  in  $(0.697, 1)$ . The picture shows that three agents with moderate  $\lambda$  assure convergence to the equilibrium despite the presence of two “turbulent” agents. Convergence is clearly not reached if no or only one agent have  $\lambda \leq 0.697$ , while an intermediate case in which agents come close to equilibrium arises when  $k = 2$ .

After this corroboration of the analytical results of the previous section, it is still to determine if our results in the “mean”  $\Gamma(w)$  game hold also in the original “median”  $\Gamma(w)$  game described in Ref. [4]. Figs. 4 and 5 are the counterparts of Figs. 1 and 3 when the median is used for computing payoffs, and rule (3) is adopted to model agents’ behavior.

The convergence properties of the mean and median cases are quite similar for the vast majority of values of  $l$  and  $k$ , thus justifying the choice to derive analytical results for the mean game to guess the behavior of the median one. However, it should be noted that some border choice of parameters bears some differences, see

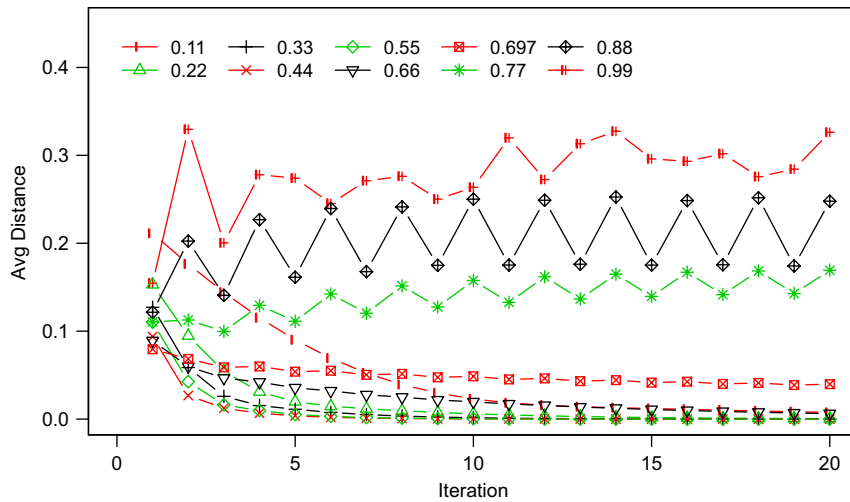


Fig. 4. Average distance from the interior equilibrium for various values of  $l$ , when  $\lambda$  is randomly sampled in  $(0, l)$  and the median is used for computing payoffs. The mean of 100 trials is depicted.

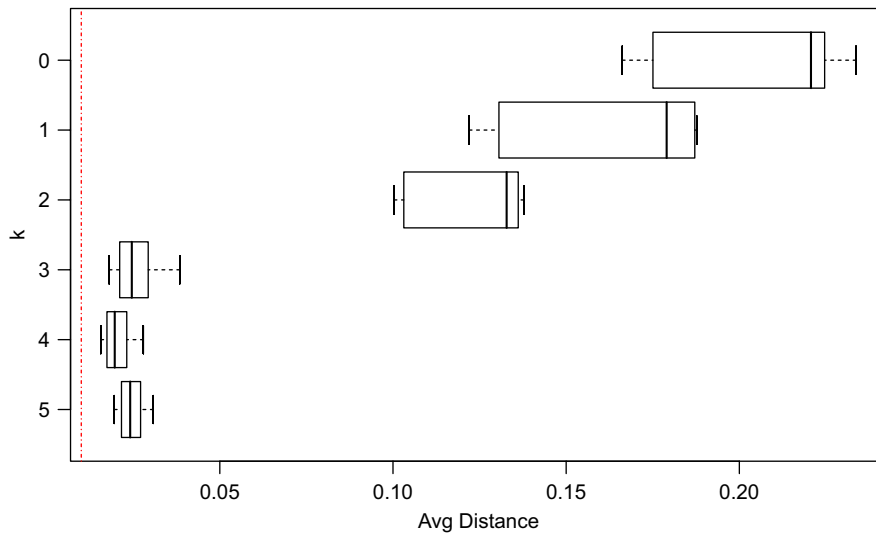


Fig. 5. Boxplots of distribution of average distances from the interior equilibrium for  $10 \leq l \leq 20$ , when  $k$  agents have  $\lambda \in (0, 0.697)$  and the median is used for computing payoffs.

for example the behavior for  $l = 0.77$  in Figs. 1 and 4. Generally speaking, convergence, if any, in the median game is more difficult to achieve than in the mean game.

We use van Huyck et al. laboratory data as a benchmark for a rough evaluation of the plausibility of convergence speed of our learning rule. In their experiments, human subjects always converge in less than 15 periods. Very good agreement with such benchmark can be obtained on average (in the simulated median game) when  $k = 3, 4, 5$  or sampling  $\lambda$  in  $[0, l]$  with  $0.22 \leq l \leq 0.66$ .

### 5. Conclusion

Motivated by the work in Ref. [4], where a representative selection dynamics for a coordination “median” game was studied, we propose a modification of the payoff function that allows a formal analysis of the game

where the median action is replaced by the mean action. The stability of the interior equilibrium is assessed in a genuine multi-agent setting, even when a considerable amount of heterogeneity is injected in the players. We assume, rather naturally, that each agent adjust his move in the direction of the best response, with different speeds of learning  $\lambda_i$ . Our selection dynamics generalizes Cournot myopic dynamics (if all  $\lambda_i = 1$ ) and closely resembles the Lucas' dynamics. This behavior is consistent with the LDT by Selten and indeed statistical tests show that the experimental data of van Huyck et al. are compatible in most cases with this theory, thus making our assumption rather plausible at the micro level.

We show by simulation that the stability result we proved is robust to the introduction of heterogeneity and even when some agents strongly overreact while playing. Finally, we show by simulation that the analytical results of our “mean” game can be transferred to the original “median” game, for which analytical results with multiple agents cannot be derived. Moreover, the convergence obtained in the micro-simulated game is stable (in both versions) and occurs at about the same speed than the experimental data for a wide set of parameters.

### Acknowledgements

Pellizzari Paolo acknowledges the support of research Grant MIUR 2004131501. Warglien acknowledges the support of research Grants MIUR 2005139342 and FIRB RBMR03A9A7. We thank M. LiCalzi and three anonymous referees for useful suggestions. All remaining errors are ours.

### Appendix

*Proof of Theorem 1.* We assume  $n$  agents play to the  $\Gamma'(w)$  game, using the learning rule (3), with  $\lambda_i = \lambda > 0$  for all  $i$ . Consider the dynamical system  $f : [0, 1]^n \rightarrow [0, 1]^n$  defined as

$$e_{i,t+1} = f_i(e_{1,t}, \dots, e_{n,t}) = (1 - \lambda)e_{i,t} + \frac{\lambda w \sum_{j=1}^n e_{j,t} [1 - \sum_{k=1}^n e_{k,t}/n]}{n}, \quad \forall i. \quad (4)$$

All fixpoints are symmetric. In fact, assume by contradiction that there is an asymmetric equilibrium such that  $e_j \neq e_k$ ; then we have

$$e_j = (1 - \lambda)e_j + \lambda w \hat{e}(1 - \hat{e}), \quad e_k = (1 - \lambda)e_k + \lambda w \hat{e}(1 - \hat{e}),$$

where  $\hat{e} = \sum_i e_i/n$ . Subtracting the two previous equations, it is easy to see that  $e_j - e_k \neq 0$  iff  $\lambda = 0$ , which is ruled out by assumption.

The fixpoints of the dynamical system  $f$  can be determined solving the equation:

$$e = \frac{w \sum e [1 - \sum e/n]}{n},$$

which is derived from (4) after simplification and omission of indexes to simplify notation. Some trivial algebra leads to the equation

$$e = we(1 - e),$$

whose solutions are 0 and  $1 - 1/w$ . Hence the two  $n$ -dimensional fixpoints of the average game are

$$(0, 0, \dots, 0)[\text{corner}], \quad \left(1 - \frac{1}{w}, 1 - \frac{1}{w}, \dots, 1 - \frac{1}{w}\right)[\text{interior}].$$

The stability properties of the fixpoints depend on the eigenvalues of the Jacobian matrix  $\mathbf{J} = [\partial f_i / \partial e_j]$ . With the position

$$K = \frac{\lambda w}{n} - \frac{2\lambda w \sum e_i}{n^2},$$



the Jacobian matrix can be written as

$$\mathbf{J} = \begin{pmatrix} 1 - \lambda + K & K & \dots & K \\ K & 1 - \lambda + K & \dots & K \\ \dots & \dots & \dots & \dots \\ K & K & \dots & 1 - \lambda + K \end{pmatrix}.$$

The eigenvalues are the zeros of the equation  $|\mathbf{J} - r\mathbf{I}| = 0$ . The determinant can be computed using appropriate row and column reductions, yielding

$$(1 - \lambda + nK - r)(1 - \lambda - r)^{n-1} = 0.$$

The first  $n - 1$  eigenvalues are  $1 - \lambda$  while the last one is

$$r = 1 - \lambda + \lambda w - \frac{2\lambda w \sum e_i}{n}.$$

As in equilibrium either  $e_i = 0$  or  $e_i = 1 - 1/w$ , for all  $i$ , it is immediate to check that the corner equilibrium is unstable and the interior equilibrium is stable<sup>4</sup> if  $0 < \lambda < 2/(w - 1)$ .  $\square$

## References

- [1] D. Fudenberg, D.K. Levine, *The Theory of Learning in Games*, The MIT Press, Cambridge, MA, 1998.
- [2] R. Marimon, Learning from learning in Economics, in: D.M. Kreps, K.F. Wallis (Eds.), *Advances in Economics and Econometrics: Theory and Applications*, vol. 1, Cambridge University Press, Cambridge (UK), pp. 278–315.
- [3] C. Camerer, *Behavioral Game Theory*, Princeton University Press, Princeton, NJ, 2003.
- [4] J. van Huyck, J. Cook, R. Battalio, Selection dynamics, asymptotic stability, and adaptive behavior, *J. Polit. Econ.* 102 (1994) 975–1005.
- [5] J.C. Harsanyi, R. Selten, *A General Theory of Equilibrium Selection in Games*, The MIT Press, Cambridge, MA, 1988.
- [6] A.P. Kirman, Whom or what does the representative individual represent?, *J. Econ. Perspect.* 6 (2) (1992) 117–136.
- [7] S. Chen, J. Duffy, C. Yeh, Equilibrium selection via adaptation: using genetic programming to model learning in a coordination game, *Electron. J. Evol. Modeling Econ. Dyn.* (2002), 1002, (<http://www.e-jemed.org/1002/index.php>).
- [8] R. Selten, R. Stocker, End behavior in sequences of finite Prisoner's dilemma supergames: a learning theory approach, *J. Econ. Behav. Organ.* 7 (1986) 47–80.
- [9] R. Selten, K. Abbink, R. Cox, Learning direction theory and the winners curse, *Exp. Econ.* 8 (2005) 5–20.
- [10] D. Lieberman, *Learning: Behavior and Cognition*, Wadsworth, Belmont, CA, 1993.
- [11] S. Selten, L. Allan, The widespread influence of the Rescorla–Wagner model, *Psychon. Bull. Rev.* 3 (s) (1996) 314–321.
- [12] E. Darmon, R. Waldeck, Convergence of reinforcement learning to Nash equilibrium: a search market experiment, *Physica A* 355 (2005) 119–130.

<sup>4</sup>Stability holds for the corner when  $|1 - \lambda| < 1$  and  $|1 - \lambda + \lambda w| < 1$ . This in turn implies that  $0 < \lambda < 2$  together with  $-2/(w - 1) < \lambda < 0$ , which is impossible. For the interior equilibrium, we must have  $0 < \lambda < 2$  and  $0 < \lambda < 2/(w - 1)$ .