Efficient Evolutionary Dynamics with Extensive–Form Games

Nicola Gatti and **Fabio Panozzo** and **Marcello Restelli**
Politecnico di Milano
Piazza Leonardo da Vinci 32
Milano, Italy
{ngatti, panozzo, restelli}@elet.polimi.it

Abstract

Evolutionary game theory combines game theory and dynamical systems and is customarily adopted to describe evolutionary dynamics in multi–agent systems. In particular, it has been proven to be a successful tool to describe multi–agent learning dynamics. To the best of our knowledge, we provide in this paper the first replicator dynamics applicable to the *sequence form* of an extensive–form game, allowing an exponential reduction of time and space w.r.t. the currently adopted replicator dynamics for normal form. Furthermore, our replicator dynamics is realization equivalent to the standard replicator dynamics for normal form. We prove our results for both discrete–time and continuous–time cases. Finally, we extend standard tools to study the stability of a strategy profile to our replicator dynamics.

Introduction

Game theory provides the most elegant tools to model strategic interaction situations among rational agents. These situations are customarily modeled as *games* (Fudenberg and Tirole 1991) in which the *mechanism* describes the rules and *strategies* describe the behavior of the agents. Furthermore, game theory provides a number of *solution concepts*. The central one is *Nash equilibrium*. Game theory assumes agents to be rational and describes "static" equilibrium states. Evolutionary game theory (Cressman 2003) drops the assumption of rationality and assumes agents to be adaptive in the attempt to describe dynamics of evolving populations. Interestingly, there are strict relations between game theory solution concepts and evolutionary game theory steady states, e.g., Nash equilibria are steady states. Evolutionary game theory is commonly adopted to study economic evolving populations (Cai, Niu, and Parsons 2007) and artificial multi–agent systems, e.g., for describing multi–agent learning dynamics (Tuyls, Hoen, and Vanschoenwinkel 2006; Tuyls and Parsons 2007; Panait, Tuyls, and Luke 2008) and as heuristics in algorithms (Kiekintveld, Marecki, and Tambe 2011). In this paper, we develop efficient techniques for evolutionary dynamics with extensive–form games.

Extensive–form games are a very important class of games. They provide a richer representation than strategic–form games, the sequential structure of decision–making being described explicitly and each agent being allowed to be free to change her mind as events unfold. The study of extensive–form games is carried out by translating the game by means of tabular representations (Shoham and Leyton-Brown 2008). The most common is the *normal form*. Its advantage is that all the techniques applicable to strategic–form games can be adopted also with this representation. However, the size of normal form grows exponentially with the size of the game tree, thus being impractical. The *agent form* is an alternative representation whose size is linear in the size of the game tree, but it makes, even with two agents, each agent's best–response problem highly non–linear. To circumvent these issues, *sequence form* was proposed (von Stengel 1996). This form is linear in the size of the game tree and does not introduce non–linearities in the best–response problem. On the other hand, standard techniques for strategic–form games cannot be adopted with such representation, e.g. (Lemke and Howson 1964), thus requiring alternative *ad hoc* techniques, e.g. (Lemke 1978). In addition, sequence form is more expressive than normal form. For instance, working with sequence form it is possible to find Nash–equilibrium refinements for extensive–form games—perfection based Nash equilibria and sequential equilibrium (Miltersen and Sørensen 2010; Gatti and Iuliano 2011)—while it is not possible with normal form.

To the best of our knowledge, there is no result dealing with the adoption of evolutionary game theory tools with sequence form for the study of extensive–form games, all the known results working with the normal form (Cressman 2003). In this paper, we originally explore this topic, providing the following main contributions.

- We show that the standard replicator dynamics for normal form cannot be adopted with the sequence form, the strategies produced by replication not being well–defined sequence–form strategies.
- We design an *ad hoc* version of the discrete–time replicator dynamics for sequence form and we show that it is sound, the strategies produced by replication being well–defined sequence–form strategies.
- We show that our replicator dynamics is realization equivalent to the standard discrete–time replicator dynamics for normal form and therefore that the two replicator dynamics evolve in the same way.
- We extend our discrete–time replicator dynamics to the

continuous–time case, showing that the same properties are satisfied and extending standard tools to study the stability of the strategies to our replicator.

Game theoretical preliminaries

Extensive–form game definition. A *perfect–information* extensive–form game (Fudenberg and Tirole 1991) is a tuple $(N, A, V, T, \iota, \rho, \chi, \mathbf{u})$, where: N is the set of agents ($i \in N$ denotes a generic agent), A is the set of actions ($A_i \subseteq A$ denotes the set of actions of agent i and $a \in A$ denotes a generic action), V is the set of decision nodes ($V_i \subseteq V$ denotes the set of decision nodes of i), T is the set of terminal nodes ($w \in V \cup T$ denotes a generic node and w_0 is root node), $\iota : V \to N$ returns the agent that acts at a given decision node, $\rho : V \to \wp(A)$ returns the actions available to agent $\iota(w)$ at w, $\chi : V \times A \to V \cup T$ assigns the next (decision or terminal) node to each pair $\langle w, a\rangle$ where a is available at w, and $\mathbf{u} = (u_1, \dots, u_{|N|})$ is the set of agents' utility functions $u_i : T \to \mathbb{R}$. Games with *imperfect information* extend those with perfect information, allowing one to capture situations in which some agents cannot observe some actions undertaken by other agents. We denote by $V_{i,h}$ the h–th *information set* of agent i. An information set is a set of decision nodes such that when an agent plays at one of such nodes she cannot distinguish the node in which she is playing. For the sake of simplicity, we assume that every information set has a different index h, thus we can univocally identify an information set by h. Furthermore, since the available actions at all nodes w belonging to the same information set h are the same, with abuse of notation, we write $\rho(h)$ in place of $\rho(w)$ with $w \in V_{i,h}$. An imperfect–information game is a tuple $(N, A, V, T, \iota, \rho, \chi, \mathbf{u}, H)$ where $(N, A, V, T, \iota, \rho, \chi, \mathbf{u})$ is a perfect–information game and $H = (H_1, \dots, H_{|N|})$ induces a partition $V_i = \bigcup_{h \in H_i} V_{i,h}$ such that for all $w, w' \in V_{i,h}$ we have $\rho(w) = \rho(w')$. We focus on games with *perfect recall* where each agent recalls all the own previous actions and the ones of the opponents (Fudenberg and Tirole 1991).

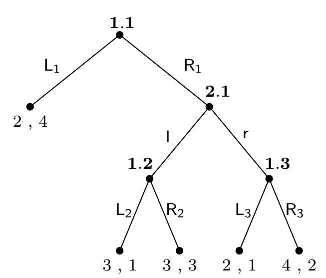

Figure 1: Example of two–agent perfect–information extensive–form game, $\mathbf{x}.\mathbf{y}$ denote the $\mathbf{y}$–th node of agent $\mathbf{x}$.

(Reduced) Normal form (von Neumann and Morgenstern 1944). It is a tabular representation in which each normal–form action, called *plan* and denoted by $p \in P_i$ where P_i is the set of plans of agent i, specifies one action a per information set. We denote by $\boldsymbol{\pi}_i$ a normal–form strategy of agent i and by $\pi_i(p)$ the probability associated with plan p. The number of plans (and therefore the size of the normal form) is exponential in the size of the game tree. The *reduced normal form* is obtained from the normal form by deleting replicated strategies (Vermeulen and Jansen 1998). Although reduced normal form can be much smaller than normal form, it is exponential in the size of the game tree.

Example 1 *The reduced normal form of the game in Fig. 1 and a pair of normal–form strategies are:*

agent 1 \ agent 2	l	r
L_1*	2, 4	2, 4
$R_1L_2L_3$	3, 1	2, 1
$R_1L_2R_3$	3, 1	4, 2
$R_1R_2L_3$	3, 3	2, 1
$R_1R_2R_3$	3, 3	4, 2

$$\pi_1 = \begin{cases} \pi_{1,L_1*} & = \frac{1}{3} \\ \pi_{1,R_1L_2L_3} & = 0 \\ \pi_{1,R_1L_2R_3} & = \frac{1}{3} \\ \pi_{1,R_1R_2L_3} & = 0 \\ \pi_{1,R_1R_2R_3} & = \frac{1}{3} \end{cases} \quad \pi_2 = \begin{cases} \pi_{2,l} & = 1 \\ \pi_{2,r} & = 0 \end{cases}$$

Agent form (Kuhn 1950; Selten 1975). It is a tabular representation in which each agent is replicated in a number of fictitious agents, each per information set, and all the fictitious agents of the same agent have the same utility. A strategy is commonly said *behavioral* and denoted by $\boldsymbol{\sigma}_i$. We denote by $\sigma_i(a)$ the probability associated with action $a \in A_i$. The agent form is linear in the size of the game tree.

Sequence form (von Stengel 1996). It is a representation constituted by a tabular and a set of constraints. Sequence–form actions are called *sequences*. A sequence $q \in Q_i$ of agent i is a set of consecutive actions $a \in A_i$ where $Q_i \subseteq Q$ is the set of sequences of agent i and Q is the set of all the sequences. A sequence can be *terminal*, if, combined with some sequence of the opponents, it leads to a terminal node, or *non–terminal* otherwise. In addition, the initial sequence of every agent, denoted by $\mathsf{q}_\varnothing$, is said *empty sequence* and, given sequence $q \in Q_i$ leading to some information set $h \in H_i$, we say that q' *extends* q and we denote by $q' = q|a$ if the last action of q' (denoted by $a(q') = a'$) is some action $a \in \rho(h)$ and q leads to h. We denote by $w = h(q)$ the node w with $a(q) \in \rho(w)$; by $q' \subseteq q$ a subsequence of q; by $\mathbf{x}_i$ the sequence–form strategy of agent i and by $x_i(q)$ the probability associated with sequence $q \in Q_i$. Finally, condition $q \to h$ is true if sequence q crosses information set h. Well–defined strategies are such that, for every information set $h \in H_i$, the probability $x_i(q)$ assigned to the sequence q leading to h is equal to the sum of the probabilities $x_i(q')$s where q' extends q at h. Sequence form constraints are $x_i(\mathsf{q}_\varnothing) = 1$ and $x_i(q) = \sum_{a \in \rho(w)} x_i(q|a)$ for every sequence q, action a, node w such that $w = h(q|a)$, and for every agent i. The agent i's utility is represented as a sparse multi–dimensional array, denoted, with an abuse of notation, by U_i, specifying the value associated with every combination of terminal sequences of all the agents. The size of the sequence form is linear in the size of the game tree.

Example 2 *The sequence form of the game in Fig. 1 and a pair of sequence–form strategies are:*

		agent 2		
		$q_\varnothing$	l	r
agent 1	$q_\varnothing$			
	L_1	2, 4		
	R_1			
	R_1L_2		3, 1	
	R_1R_2		3, 3	
	R_1L_3			2, 1
	R_1R_3			4, 2

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ \frac{1}{3} \\ \frac{2}{3} \\ \frac{1}{3} \\ \frac{1}{3} \\ 0 \\ \frac{2}{3} \end{bmatrix} \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

Replicator dynamics. The standard discrete–time replicator equation with two agents is (Cressman 2003):

$$\pi_1(p, t+1) = \pi_1(p, t) \cdot \frac{\mathbf{e}_p^T \cdot U_1 \cdot \boldsymbol{\pi}_2(t)}{\boldsymbol{\pi}_1^T(t) \cdot U_1 \cdot \boldsymbol{\pi}_2(t)} \tag{1}$$

$$\pi_2(p, t+1) = \pi_2(p, t) \cdot \frac{\boldsymbol{\pi}_1^T(t) \cdot U_2 \cdot \mathbf{e}_p}{\boldsymbol{\pi}_1^T(t) \cdot U_2 \cdot \boldsymbol{\pi}_2(t)} \tag{2}$$

while the continuous–time one is

$$\dot{\pi}_1(p) = \pi_1(p) \cdot [(\mathbf{e}_p - \boldsymbol{\pi}_1)^T \cdot U_1 \cdot \boldsymbol{\pi}_2] \tag{3}$$

$$\dot{\pi}_2(p) = \pi_2(p) \cdot [\boldsymbol{\pi}_1^T \cdot U_2 \cdot (\mathbf{e}_p - \boldsymbol{\pi}_2)] \tag{4}$$

where $\mathbf{e}_p$ is the vector in which the p–th component is "1" and the others are "0".

Discrete–time replicator dynamics for sequence–form representation

Initially, we show that the standard discrete–time replicator dynamics for normal form cannot be directly applied when sequence form is adopted. Standard replicator dynamics applied to the sequence form is easily obtained by considering each sequence q as a plan p and thus substituting $\mathbf{e}_q$ to $\mathbf{e}_p$ in (1)–(2) where $\mathbf{e}_q$ is zero for all the components q' such that $q' \neq q$ and one for the component q' such that $q' = q$.

Proposition 3 *The replicator (1)–(2) does not satisfy the sequence–form constraints.*

Proof. The proof is by counterexample. Consider $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ equal to the strategies used in Example 2. At time $t+1$ the strategy profile generated by (1)–(2) is:

$$\mathbf{x}_1^T(t+1) = \begin{bmatrix} 0 & \frac{1}{3} & 0 & \frac{1}{2} & \frac{1}{6} & 0 & 0 \end{bmatrix} \quad \mathbf{x}_2^T(t+1) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

that does not satisfy the sequence–form constraints, e.g., $x_i(q_\varnothing, t+1) \neq 1$ for all i. □

The critical issue behind the failure of the standard replicator dynamics lies in the definition of vector $\mathbf{e}_q$. Now we describe how the standard discrete–time replicator dynamics can be modified to be applied to the sequence form. In our variation, we substitute $\mathbf{e}_q$ with an opportune vector $\mathbf{g}_q$ that depends on the strategy $\mathbf{x}_i(t)$ and it is generated as described in Algorithm 1, obtaining:

$$x_1(q, t+1) = x_1(q, t) \cdot \frac{\mathbf{g}_q^T(\mathbf{x}_1(t)) \cdot U_1 \cdot \mathbf{x}_2(t)}{\mathbf{x}_1^T(t) \cdot U_1 \cdot \mathbf{x}_2(t)} \tag{5}$$

$$x_2(q, t+1) = x_2(q, t) \cdot \frac{\mathbf{x}_1^T(t) \cdot U_2 \cdot \mathbf{g}_q(\mathbf{x}_2(t))}{\mathbf{x}_1^T(t) \cdot U_2 \cdot \mathbf{x}_2(t)} \tag{6}$$

The basic idea behind the construction of vector g_q is:

Algorithm 1 generate_$\mathbf{g}_q(\mathbf{x}_i(t))$

```
g_q(x_i(t)) = 0
if x_i(q,t) ≠ 0 then
    for q' ∈ Q_i s.t. q' ⊆ q do
        g_q(q', x_i(t)) = 1
        for q'' ∈ Q_i s.t. q'' ∩ q = q' and q'' = q'|a|... : a ∈ ρ(h), q ↛ h do
            g_q(q'', x_i(t)) = x_i(q'',t) / x_i(q',t)
return g_q(x_i(t))
```

- assigning "1" to the probability of all the sequences contained in q,
- normalizing the probability of the sequences extending the contained in q,
- assigning "0" to the probability of all the other sequences.

We describe the generation of vector $\mathbf{g}_q(\mathbf{x}_i(t))$, for clarity we use as running example the generation of $\mathbf{g}_{R_1R_3}(\mathbf{x}_1(t))$ related to Example 2:

- all the components of $\mathbf{g}_q(\mathbf{x}_i(t))$ are initialized equal to "0", e.g.,

$$\mathbf{g}_{R_1R_3}(\mathbf{x}_1(t))^T = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

- if sequence q is played, the algorithm assigns:
  - "1" to all the components $g_q(q', \mathbf{x}_i(t))$ of $\mathbf{g}_q(\mathbf{x}_i(t))$ where $q' \subseteq q$ (i.e., q' is a subsequence of q), e.g.,

$$\mathbf{g}_{R_1R_3}(\mathbf{x}_1(t))^T = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

  - "$\frac{x_i(q'',t)}{x_i(q',t)}$" to all the components $g_q(q'', \mathbf{x}_i(t))$ of $\mathbf{g}_q(\mathbf{x}_i(t))$ where $q' \subseteq q$ with $q' = q'' \cap q$ and sequence q'' is defined as $q'' = q'|a|\dots$ with $a \in \rho(h)$ and $q \not\rightarrow h$ (i.e., q' is a subsequence of q and q'' extends q' off the path identified by q), e.g.,

$$\mathbf{g}_{R_1R_3}(\mathbf{x}_1(t))^T = \begin{bmatrix} 1 & 0 & 1 & \frac{1}{2} & \frac{1}{2} & 0 & 1 \end{bmatrix}$$

  - all the other components are left equal to "0",
- if sequence q is not played, $\mathbf{g}_q(\mathbf{x}_i(t))$ can be arbitrary, since the q–th equation of (5)–(6) is always zero given that $x_i(q,t) = 0$ for every t.

All the vectors $\mathbf{g}_q(\mathbf{x}_1(t))$ of Example 2 are:

	$\mathbf{g}_{q_\varnothing}$	$\mathbf{g}_{L1}$	$\mathbf{g}_{R_1}$	$\mathbf{g}_{R_1L_2}$	$\mathbf{g}_{R_1R_2}$	$\mathbf{g}_{R_1L_3}$	$\mathbf{g}_{R_1R_3}$
$q_\varnothing$	1	1	1	1	1	1	1
L_1	$\frac{1}{3}$	1	0	0	0	0	0
R_1	$\frac{2}{3}$	0	1	1	1	1	1
R_1L_2	$\frac{1}{3}$	0	$\frac{1}{2}$	1	0	$\frac{1}{2}$	$\frac{1}{2}$
R_1R_2	$\frac{1}{3}$	0	$\frac{1}{2}$	0	1	$\frac{1}{2}$	$\frac{1}{2}$
R_1L_3	0	0	0	0	0	1	0
R_1R_3	$\frac{2}{3}$	0	1	1	1	0	1

We show that replicator dynamics (5)–(6) do not violate sequence–form constraints.

Theorem 4 *Given a well–defined sequence–form strategy profile $(\mathbf{x}_1(t), \mathbf{x}_2(t))$, the output strategy profile $(\mathbf{x}_1(t+1), \mathbf{x}_2(t+1))$ of replicator dynamics (5)–(6) satisfies sequence–form constraints.*

Proof. The constraints forced by sequence form are:

- $x_i(\mathsf{q}_\varnothing, t) = 1$ for every i,
- $x_i(q,t) = \sum_{a \in \rho(w)} x_i(q|a,t)$ for every sequence q, action a, node w such that $w = h(q|a)$, and for every agent i.

Assume, by hypothesis of the theorem, that the above constraints are satisfied at t, we need to prove that constraints

$$x_i(\mathsf{q}_\varnothing, t+1) = 1 \tag{7}$$

$$x_i(q, t+1) = \sum_{a \in \rho(w)} x_i(q|a, t+1) \tag{8}$$

are satisfied. Constraint (7) always holds because $\mathbf{g}_{\mathsf{q}_\varnothing}(\mathbf{x}_1(t)) = \mathbf{x}_1(t)$. We rewrite constraints (8) as

$$x_i(q,t) \cdot \frac{\mathbf{g}_q^T(\mathbf{x}_i(t)) \cdot U_i \cdot \mathbf{x}_{-i}(t)}{\mathbf{x}_i^T(t) \cdot U_i \cdot \mathbf{x}_{-i}(t)} = \\ = \sum_{a \in \rho(w)} \left(x_i(q|a,t) \cdot \frac{\mathbf{g}_{q|a}^T(\mathbf{x}_i(t)) \cdot U_i \cdot \mathbf{x}_{-i}(t)}{\mathbf{x}_i^T(t) \cdot U_i \cdot \mathbf{x}_{-i}(t)} \right) \tag{9}$$

Conditions (9) hold if the following condition holds

$$x_i(q,t) \cdot \mathbf{g}_q^T(\mathbf{x}_i(t)) = \sum_{a \in \rho(w)} \left(x_i(q|a,t) \cdot \mathbf{g}_{q|a}^T(\mathbf{x}_i(t)) \right) \tag{10}$$

Notice that condition (10) is a vector of equalities, one per sequence q'. Condition (10) is trivially satisfied for components q' such that $g_q(q', \mathbf{x}_i(t)) = 0$. To prove the condition for all the other components, we introduce two lemmas.

Lemma 5 *Constraint (10) holds for all components $g_q(q', \mathbf{x}_i(t))$ of $\mathbf{g}_q(\mathbf{x}_i(t))$ such that $q' \subseteq q$.*

Proof. By construction, $g_q(q', \mathbf{x}_i(t)) = 1$ for every $q' \subseteq q$. For every extension $q|a$ of q, we have that $q' \subseteq q \subset q|a$. For this reason $g_{q|a}(q', \mathbf{x}_i(t)) = 1$. Thus

$$x_i(q,t) \cdot g_q(q', \mathbf{x}_i(t)) = \sum_{a \in \rho(w)} \left(x_i(q|a,t) \cdot g_{q|a}(q', \mathbf{x}_i(t)) \right) \quad \text{iff}$$

$$x_i(q,t) \cdot 1 = \sum_{a \in \rho(w)} x_i(q|a,t) \cdot 1$$

that holds by hypothesis. Therefore the lemma is proved. □

Lemma 6 *Constraint (10) holds for all components $g_q(q'', \mathbf{x}_i(t))$ of $\mathbf{g}_q(\mathbf{x}_i(t))$ where $q' \subseteq q$ with $q' = q'' \cap q$ and sequence $q'' = q'|a'|\dots$ with $a' \in \rho(h)$ and $q \not\to h$.*

Proof. For all q'', $g_q(q'', \mathbf{x}_i(t)) = \frac{x_i(q'',t)}{x_i(q',t)}$ by construction. In the right side term of (10), for all a we can have either $q|a \not\subset q''$ or $q|a \subset q''$. In the former we have that $g_{q|a}(q'', \mathbf{x}_i(t)) = \frac{x_i(q'',t)}{x_i(q',t)}$, in the latter there exists only one action a such that $g_{q|a}(q'', \mathbf{x}_i(t)) = \frac{x_i(q'',t)}{x_i(q|a,t)}$, while for the other actions a^* the value of $g_{q|a^*}(q'', \mathbf{x}_i(t))$ is zero. Hence, we can have two cases: if $q|a \not\subset q''$, then

$$x_i(q,t) \cdot g_q(q'', \mathbf{x}_i(t)) = \sum_{a \in \rho(w)} \left(x_i(q|a,t) \cdot g_{q|a}(q'', \mathbf{x}_i(t)) \right) \quad \text{iff}$$

$$x_i(q,t) \cdot \frac{x_i(q'',t)}{x_i(q',t)} = \sum_{a \in \rho(w)} \left(x_i(q|a,t) \cdot \frac{x_i(q'',t)}{x_i(q',t)} \right)$$

that holds by hypothesis, otherwise if $q|a \subset q''$, then

$$x_i(q,t) \cdot g_q(q'', \mathbf{x}_i(t)) = \sum_{a \in \rho(w)} \left(x_i(q|a,t) \cdot g_{q|a}(q'', \mathbf{x}_i(t)) \right) \quad \text{iff}$$

$$x_i(q,t) \cdot \frac{x_i(q'',t)}{x_i(q,t)} = x_i(q|a,t) \cdot \frac{x_i(q'',t)}{x_i(q|a,t)}$$

that always holds. Therefore the lemma is proved. □

From the application of Lemmas 5 and 6, it follows that condition (10) holds. □

Replicator dynamics realization equivalence

There is a well–known relation, based on the concept of *realization*, between normal–form and sequence–form strategies. In order to exploit it, we introduce two results from (Koller, Megiddo, and von Stengel 1996).

Definition 7 (Realization equivalent) *Two strategies of an agent are* realization equivalent *if, for any fixed strategies of the other agents, both strategies define the same probabilities for reaching the nodes of the game tree.*

Proposition 8 *For an agent with perfect recall, any normal–form strategy is realization equivalent to a sequence–form strategy.*

We recall in addition that each pure sequence–form strategy corresponds to a pure normal–form strategy in the reduced normal form (Koller, Megiddo, and von Stengel 1996). We can show that the evolutionary dynamics of (5)–(6) are realization equivalent to the evolutionary dynamics of the normal–form replicator dynamics and therefore that the two replicator dynamics evolve in the same way.

Initially, we introduce the following lemma that we will exploit to prove the main result.

Lemma 9 *Given*

- *a reduced–normal–form strategy $\boldsymbol{\pi}_i(t)$ of agent i,*
- *a sequence–form strategy $\mathbf{x}_i(t)$ realization equivalent to $\boldsymbol{\pi}_i(t)$,*

it holds that $x_i(q|a,t) \cdot \mathbf{g}_{q|a}^T(\mathbf{x}_i(t))$ is realization equivalent to $\sum_{p \in P: a \in p} \left(\pi_i(p,t) \cdot \mathbf{e}_p^T \right)$ for all $a \in A_i$ and $q \in Q_i$ with $q|a \in Q_i$.

Proof. We denote by $\tilde{\mathbf{x}}_p(t)$ the sequence–form strategy realization equivalent to $\mathbf{e}_p(t)$. According to (Koller, Megiddo, and von Stengel 1996), we can rewrite the thesis of the theorem as

$$x_i(q|a,t) \cdot \mathbf{g}_{q|a}^T(\mathbf{x}_i(t)) = \sum_{p \in P: a \in p} \left(\pi_i(p,t) \cdot \tilde{\mathbf{x}}_p(t)^T \right) \quad \forall a \in A_i \tag{11}$$

Notice that, for each action a and sequence q such that $q|a \in Q_i$, condition (11) is a vector of equality conditions. Given a and q, two cases are possible:

1. $x_i(q|a,t) = 0$ and then $\sum_{p \in P: a \in p} \pi_i(p,t) = 0$, thus conditions (11) hold;
2. $x_i(q|a,t) \neq 0$, in this case:

- for all components $g_{q|a}(q', \mathbf{x}_i(t))$ of $\mathbf{g}_{q|a}(\mathbf{x}_i(t))$ and $\tilde{x}_p(q', t)$ of $\tilde{\mathbf{x}}_p(t)$ such that $q' \subseteq q|a$, we have that $\tilde{x}_p(q', t) = 1$ for all $p \in P$ with $a \in p$ and Algorithm 1 sets $g_{q|a}(q', \mathbf{x}_i(t)) = 1$, thus we can rewrite (11) as

$$x_i(q|a, t) \cdot g_{q|a}(q', \mathbf{x}_i(t)) = \sum_{p \in P: a \in p} \left(\pi_i(p, t) \cdot \tilde{x}_p(q', t)\right) \qquad \text{iff}$$

$$x_i(q|a, t) \cdot 1 = \sum_{p \in P: a \in p} \left(\pi_i(p, t) \cdot 1\right)$$

  that holds by hypothesis and thus conditions (11) hold;
- for all components $g_{q|a}(q'', \mathbf{x}_i(t))$ of $\mathbf{g}_{q|a}(\mathbf{x}_i(t))$ and $\tilde{x}_p(q'', t)$ of $\tilde{\mathbf{x}}_p(t)$ such that q'' such that $q'' \cap q = q'$ and sequence $q'' = q'|a'|\dots$ with $a' \in \rho(h)$ and $q \not\to h$, we have that $\tilde{x}_p(q'', t) = 1$ for all $p \in P$ with $a, a(q'') \in p$ and "0" otherwise, and Algorithm 1 sets $g_{q|a}(q'', \mathbf{x}_i(t)) = \frac{x_i(q'', t)}{x_i(q', t)}$, thus we can rewrite (11) as

$$x_i(q|a, t) \cdot g_{q|a}(q'', \mathbf{x}_i(t)) = \sum_{p \in P: a \in p} \left(\pi_i(p, t) \cdot \tilde{x}_p(q'', t)\right) \qquad \text{iff}$$

$$x_i(q|a, t) \cdot \frac{x_i(q'', t)}{x_i(q', t)} = \sum_{p \in P: a, a(q'') \in p} \left(\pi_i(p, t) \cdot 1\right)$$

  Using the relationship with the behavioral strategies, we can write

$$x_i(q|a, t) \cdot \frac{x_i(q'', t)}{x_i(q', t)} = \prod_{a' \in q|a} \sigma_i(a', t) \cdot \frac{\prod_{a' \in q''} \sigma_i(a', t)}{\prod_{a' \in q'} \sigma_i(a', t)}$$

  Being $q' \subseteq q|a$ and $q' \subseteq q''$ we have

$$x_i(q|a, t) \cdot \frac{x_i(q'', t)}{x_i(q', t)} = \prod_{a' \in q|a \setminus q'} \sigma_i(a', t) \cdot \prod_{a' \in q'} \sigma_i(a', t) \cdot \prod_{a' \in q'' \setminus q'} \sigma_i(a', t) = \prod_{a^* \in \bigcup_{a' \in \{a, a(q'')\}} q(a')} \sigma_i(a^*, t)$$

  that can be easily rewrite as—for details (Gatti, Panozzo, and Restelli 2013)—

$$\sum_{p \in P: a, a(q'') \in p} \pi_i(p, t) = \prod_{a^* \in \bigcup_{a' \in \{a, a(q'')\}} q(a')} \sigma_i(a^*, t)$$

  and therefore conditions (11) hold.

This completes the proof of the lemma. □

Now we state the main result. It allows us to study the evolution of a strategy in a game directly in sequence form, instead of using the normal form, and it guarantees that the two dynamics (sequence and normal) are equivalent.

Theorem 10 *Given*

- *a normal–form strategy profile* $(\boldsymbol{\pi}_1(t), \boldsymbol{\pi}_2(t))$ *and its evolution* $(\boldsymbol{\pi}_1(t+1), \boldsymbol{\pi}_2(t+1))$ *according to (1)–(2),*
- *a sequence–form strategy profile* $(\mathbf{x}_1(t), \mathbf{x}_2(t))$ *and its evolution* $(\mathbf{x}_1(t+1), \mathbf{x}_2(t+1))$ *according to (5)–(6),*

if $(\boldsymbol{\pi}_1(t), \boldsymbol{\pi}_2(t))$ *and* $(\mathbf{x}_1(t), \mathbf{x}_2(t))$ *are realization equivalent, then also* $(\boldsymbol{\pi}_1(t+1), \boldsymbol{\pi}_2(t+1))$ *and* $(\mathbf{x}_1(t+1), \mathbf{x}_2(t+1))$ *are realization equivalent.*

Proof. Assume, by hypothesis of the theorem, that $(\mathbf{x}_1(t), \mathbf{x}_2(t))$ is realization equivalent to $(\boldsymbol{\pi}_1(t), \boldsymbol{\pi}_2(t))$. Thus, according to (Koller, Megiddo, and von Stengel 1996), for every agent i it holds

$$x_i(q|a, t) = \sum_{p \in P: a \in p} \pi_i(p, t) \qquad \forall a \in A_i$$

We need to prove that the following conditions hold:

$$x_i(q|a, t+1) = \sum_{p \in P: a \in p} \pi_i(p, t+1) \qquad \forall a \in A_i \tag{12}$$

By applying the definition of replicator dynamics, we can rewrite the conditions (12) as:

$$x_i(q|a, t) \cdot \frac{\mathbf{g}^T_{q|a}(\mathbf{x}_i(t)) \cdot U_i \cdot \mathbf{x}_{-i}(t)}{\mathbf{x}^T_i(t) \cdot U_i \cdot \mathbf{x}_{-i}(t)} = \sum_{p \in P: a \in p} \left(\pi_i(p, t) \cdot \frac{\mathbf{e}^T_p \cdot U_i \cdot \boldsymbol{\pi}_{-i}(t)}{\boldsymbol{\pi}^T_i(t) \cdot U_i \cdot \boldsymbol{\pi}_{-i}(t)}\right) \qquad \forall a \in A_i \tag{13}$$

Given that, by hypothesis, $\mathbf{x}^T_i(t) \cdot U_i \cdot \mathbf{x}_{-i}(t) = \boldsymbol{\pi}^T_i(t) \cdot U_i \cdot \boldsymbol{\pi}_{-i}(t)$, we can rewrite conditions (13) as:

$$x_i(q|a, t) \cdot \mathbf{g}^T_{q|a}(\mathbf{x}_i(t)) \cdot U_i \cdot \mathbf{x}_{-i}(t) = \sum_{p \in P: a \in p} \left(\pi_i(p, t) \cdot \mathbf{e}^T_p \cdot U_i \cdot \boldsymbol{\pi}_{-i}(t)\right) \qquad \forall a \in A_i$$

These conditions hold if and only if $\sum_{p \in P: a \in p} \left(\pi_i(p, t) \cdot \mathbf{e}^T_p\right)$ is realization equivalent to $x_i(q|a, t) \cdot \mathbf{g}^T_{q|a}(\mathbf{x}_i(t))$. By Lemma 9, this equivalence holds. □

Continuous–time replicator dynamics for sequence–form representation

The sequence–form continuous–time replicator equation is

$$\dot{x}_1(q, t) = x_1(q, t) \cdot [(\mathbf{g}_q(\mathbf{x}_1(t)) - \mathbf{x}_1(t))^T \cdot U_1 \cdot \mathbf{x}_2(t)] \tag{14}$$

$$\dot{x}_2(q, t) = x_2(q, t) \cdot [\mathbf{x}_1(t)^T \cdot U_2 \cdot (\mathbf{g}_q(\mathbf{x}_2(t) - \mathbf{x}_2(t))] \tag{15}$$

Theorem 11 *Given a well–defined sequence–form strategy profile* $(\mathbf{x}_1(t), \mathbf{x}_2(t))$*, the output strategy profile* $(\mathbf{x}_1(t+\Delta t), \mathbf{x}_2(t+\Delta t))$ *of replicator dynamics (14)–(15) satisfies sequence–form constraints.*

Proof. The constraints forced by sequence form are:

- $x_i(\mathsf{q}_\varnothing, t) = 1$ for every i,
- $x_i(q, t) = \sum_{a \in \rho(w)} x_i(q|a, t)$ for every sequence q, action a, node w such that $w = h(q|a)$, and for every agent i.

Assume, by hypothesis of the theorem, that constraints are satisfied at a given time point t, we need to prove that constraints

$$x_i(\mathsf{q}_\varnothing, t+\Delta t) = 1 \tag{16}$$

$$x_i(q, t+\Delta t) = \sum_{a \in \rho(w)} x_i(q|a, t+\Delta t) \tag{17}$$

are satisfied. Constraint (16) always holds because $\mathbf{g}_q(\mathbf{x}_1(t)) = \mathbf{x}_1(t)$. We rewrite constraints (17) as

$$x_i(q, t) \cdot [(\mathbf{g}_q(\mathbf{x}_i(t)) - \mathbf{x}_i(t))^T \cdot U_i \cdot \mathbf{x}_{-i}(t)] = \sum_{a \in \rho(w)} \left(x_i(q|a) \cdot [(\mathbf{g}_{q|a}(\mathbf{x}_i(t)) - \mathbf{x}_i(t))^T \cdot U_i \cdot \mathbf{x}_{-i}(t)]\right) \tag{18}$$

Conditions (18) hold if the following conditions hold

$$x_i(q, t) \cdot \mathbf{g}^T_q(\mathbf{x}_i(t)) = \sum_{a \in \rho(w)} \left(x_i(q|a, t) \cdot \mathbf{g}^T_{q|a}(\mathbf{x}_i(t))\right) \tag{19}$$

Notice that condition (19) is a vector of equalities. The above condition is trivially satisfied for components q' such that $g_q(q', \mathbf{x}_i(t)) = 0$. From the application of Lemmas 5 and 6, the condition (19) holds also for all the other components. □

Theorem 12 *Given*

- *a normal–form strategy profile* $(\boldsymbol{\pi}_1(t), \boldsymbol{\pi}_2(t))$ *and its evolution* $(\boldsymbol{\pi}_1(t+\Delta t), \boldsymbol{\pi}_2(t+\Delta t))$ *according to (3)–(4),*
- *a sequence–form strategy profile* $(\mathbf{x}_1(t), \mathbf{x}_2(t))$ *and its evolution* $(\mathbf{x}_1(t+\Delta t), \mathbf{x}_2(t+\Delta t))$ *according to (14)–(15),*

if $(\boldsymbol{\pi}_1(t), \boldsymbol{\pi}_2(t))$ *and* $(\mathbf{x}_1(t), \mathbf{x}_2(t))$ *are realization equivalent, then also* $(\boldsymbol{\pi}_1(t+\Delta t), \boldsymbol{\pi}_2(t+\Delta t))$ *and* $(\mathbf{x}_1(t+\Delta t), \mathbf{x}_2(t+\Delta t))$ *are realization equivalent.*

Proof. Assume, by hypothesis of the theorem, that $(\mathbf{x}_1(t), \mathbf{x}_2(t))$ is realization equivalent to $(\boldsymbol{\pi}_1(t), \boldsymbol{\pi}_2(t))$. Thus, according to (Koller, Megiddo, and von Stengel 1996), for every agent i it holds

$$x_i(q|a,t) = \sum_{p \in P: a \in p} \pi_i(p,t) \qquad \forall a \in A_i$$

We need to prove that the following conditions hold:

$$x_i(q|a,t+\Delta t) = \sum_{p \in P: a \in p} \pi_i(p,t+\Delta t) \qquad \forall a \in A_i \tag{20}$$

By applying the definition of replicator dynamics, we can rewrite the conditions (20) as:

$$x_i(q|a,t) \cdot [(\mathbf{g}_{q|a}(\mathbf{x}_i(t)) - \mathbf{x}_i(t))^T \cdot U_i \cdot \mathbf{x}_{-i}(t)] = \\ = \sum_{p \in P: a \in p} \left(\pi_i(p,t) \cdot [(\mathbf{e}_p - \boldsymbol{\pi}_i(t))^T \cdot U_i \cdot \boldsymbol{\pi}_{-i}(t)] \right) \qquad \forall a \in A_i \tag{21}$$

Given that, by hypothesis, $\mathbf{x}_i^T(t) \cdot U_i \cdot \mathbf{x}_{-i}(t) = \boldsymbol{\pi}_i^T(t) \cdot U_i \cdot \boldsymbol{\pi}_{-i}(t)$, we can rewrite conditions (21) as:

$$x_i(q|a,t) \cdot \mathbf{g}_{q|a}^T(\mathbf{x}_i(t)) \cdot U_i \cdot \mathbf{x}_{-i}(t) = \\ = \sum_{p \in P: a \in p} \left(\pi_i(p,t) \cdot \mathbf{e}_p^T \cdot U_i \cdot \boldsymbol{\pi}_{-i}(t) \right) \qquad \forall a \in A_i$$

These conditions hold if and only if $\sum_{p \in P: a \in p} \left(\pi_i(p,t) \cdot \mathbf{e}_p^T \right)$ is realization equivalent to $x_i(q|a,t) \cdot \mathbf{g}_{q|a}^T(\mathbf{x}_i(t))$. By Lemma 9, this equivalence holds. □

Analyzing the stability of a strategy profile

We focus on characterizing a strategy profile in terms of evolutionary stability. When the continuous–time replicator dynamics for normal–form is adopted, evolutionary stability can be analyzed by studying the eigenvalues of the Jacobian in that point (Arrowsmith and Place 1992)—non–positiveness of the eigenvalues is a necessary condition for asymptotical stability, while strict negativeness of the eigenvalues is sufficient. The Jacobian is

$$J = \begin{bmatrix} \dfrac{\partial \dot{x}_1(q_i,t)}{\partial x_1(q_j,t)} & \dfrac{\partial \dot{x}_1(q_i,t)}{\partial x_2(q_l,t)} \\ \\ \dfrac{\partial \dot{x}_2(q_k,t)}{\partial x_1(q_j,t)} & \dfrac{\partial \dot{x}_2(q_k,t)}{\partial x_2(q_l,t)} \end{bmatrix} \qquad \begin{matrix} \forall q_i, q_j \in Q_1, \\ q_k, q_l \in Q_2 \end{matrix}$$

In order to study the Jacobian of our replicator dynamics, we need to complete the definition of $\mathbf{g}_q(\mathbf{x}_i(t))$. Indeed, we observe that some components of $\mathbf{g}_q(\mathbf{x}_i(t))$ are left arbitrary by Algorithm 1. Exactly, some q'' that are related to q' with $x_i(q',t) = 0$. While it is not necessary to assign values to such components during the evolution of the replicator dynamics, it is necessary when we study the Jacobian. The rationale follows. If $x_i(q',t) = 0$, then it will remain zero even after t. Instead, if, after the dynamics converged to a point, such a point has $x_i(q') = 0$ for some q', it might be the case that along the dynamics it holds $x_i(q') \neq 0$. Thus, in order to define these components of $\mathbf{g}_q(\mathbf{x}_i(t))$, we need to reason backward, assigning the values that they would have in the case such sequence would be played with a probability that goes to zero. In absence of degeneracy, Algorithm 2 addresses this issue assigning a value of "1" to a sequence q'' if it is the (unique, the game being non–degenerate) best response among the sequences extending q' and "0" otherwise, because at the convergence the agents play only the best response sequences. Notice that, in this case, $\mathbf{g}_q(\mathbf{x}_i(t), \mathbf{x}_{-i}(t))$ depends on both agents' strategies.

Algorithm 2 generate_$\mathbf{g}_\mathsf{q}(\mathbf{x}_i(t), \mathbf{x}_{-i}(t))$

$\mathbf{g}_q(\mathbf{x}_i(t), \mathbf{x}_{-i}(t)) = \mathbf{0}$
for $q' \in Q_i$ s.t. $q' \subseteq q$ **do**
    $g_q(q', \mathbf{x}_i(t), \mathbf{x}_{-i}(t)) = 1$
for $q'' \in Q_i$ s.t. $q'' \cap q = q'$ **and** $q'' = q'|a|\ldots : a \in \rho(h), q \not\rightarrow h$ **do**
    if $x_i(q',t) \neq 0$ **then**
        $g_q(q'', \mathbf{x}_i(t), \mathbf{x}_{-i}(t)) = \frac{x_i(q'',t)}{x_i(q',t)}$
    else if $q'' = \text{argmax}_{q^*: a(q^*) \in \rho(h)} \mathbb{E}[U_i(q^*, \mathbf{x}_{-i})]$ **then**
        $g_q(q'', \mathbf{x}_i(t), \mathbf{x}_{-i}(t)) = 1$
return $\mathbf{g}_q(\mathbf{x}_i(t), \mathbf{x}_{-i}(t))$

Given the above complete definition of $\mathbf{g}_q$, we can observe that all the components of $\mathbf{g}_q(\mathbf{x}_i(t), \mathbf{x}_{-i}(t))$ generated by Algorithm 2 are differentiable, being "0" or "1" or "$\frac{x_i(q'',t)}{x_i(q',t)}$". Therefore, we can derive the Jacobian as:

$$\frac{\partial \dot{x}_1(q_i,t)}{\partial x_1(q_j,t)} = \begin{cases} (\mathbf{g}_{q_i}(\mathbf{x}_1(t), \mathbf{x}_2(t)) - \mathbf{x}_1(t))^T \cdot U_1 \cdot \mathbf{x}_2(t) + x_1(q_i,t) \cdot \\ \cdot \left[\left(\dfrac{\partial \mathbf{g}_{q_i}(\mathbf{x}_1(t), \mathbf{x}_2(t))}{\partial x_1(q_j,t)} - \mathbf{e}_i \right)^T \cdot U_1 \cdot \mathbf{x}_2(t) \right] & \text{if } i = j \\ x_1(q_i,t) \cdot \left[\left(\dfrac{\partial \mathbf{g}_{q_i}(\mathbf{x}_1(t), \mathbf{x}_2(t))}{\partial x_1(q_j,t)} - \mathbf{e}_j \right)^T \cdot U_1 \cdot \mathbf{x}_2(t) \right] & \text{if } i \neq j \end{cases}$$

$$\frac{\partial \dot{x}_1(q_i,t)}{\partial x_2(q_l,t)} = x_1(q_i,t) \cdot [(\mathbf{g}_{q_i}(\mathbf{x}_1(t), \mathbf{x}_2(t)) - \mathbf{x}_1(t))^T \cdot U_1 \cdot \mathbf{e}_l]$$

$$\frac{\partial \dot{x}_2(q_k,t)}{\partial x_1(q_j,t)} = x_2(q_k,t) \cdot [\mathbf{e}_j^T \cdot U_2 \cdot (\mathbf{g}_{q_k}(\mathbf{x}_2(t), \mathbf{x}_1(t)) - \mathbf{x}_2(t))]$$

$$\frac{\partial \dot{x}_2(q_k,t)}{\partial x_2(q_l,t)} = \begin{cases} \mathbf{x}_1(t)^T \cdot U_2 \cdot (\mathbf{g}_{q_k}(\mathbf{x}_2(t), \mathbf{x}_1(t)) - \mathbf{x}_2(t)) + x_2(q_k,t) \cdot \\ \cdot \left[\mathbf{x}_1(t)^T \cdot U_2 \cdot \left(\dfrac{\partial \mathbf{g}_{q_k}(\mathbf{x}_2(t), \mathbf{x}_1(t))}{\partial x_2(q_l,t)} - \mathbf{e}_k \right) \right] & \text{if } k = l \\ x_2(q_k,t) \cdot \left[\mathbf{x}_1(t)^T \cdot U_2 \cdot \left(\dfrac{\partial \mathbf{g}_{q_k}(\mathbf{x}_2(t), \mathbf{x}_1(t))}{\partial x_2(q_l,t)} - \mathbf{e}_l \right) \right] & \text{if } k \neq l \end{cases}$$

With degenerate games, given a opponent's strategy profile $\mathbf{x}_{-i}(t)$ and a sequence $q \in Q_i$ such that $x_i(q,t) = 0$, we can have multiple best responses. Consider, e.g., the game in Example 2, with $\mathbf{x}_1^T(t) = [\ 1\ \ 1\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ ]$, $\mathbf{x}_2^T(t) = [\ 1\ \ 1\ \ 0\ ]$ and compute $\mathbf{g}_{\mathsf{R}_1\mathsf{L}_3}(\mathbf{x}_1(t), \mathbf{x}_2(t))$: both sequences $\mathsf{R}_1\mathsf{L}_2$ and $\mathsf{R}_1\mathsf{R}_2$ are best responses to $\mathbf{x}_2(t)$. Reasoning backward, we have different vectors $\mathbf{g}_q(\mathbf{x}_i, \mathbf{x}_{-i})$ for different dynamics. More precisely, we can partition the strategy space around $(\mathbf{x}_i, \mathbf{x}_{-i})$, associating a different best response with a different subspace and therefore with a different $\mathbf{g}_q(\mathbf{x}_i, \mathbf{x}_{-i})$. Thus, in principle, in order to study the stability of a strategy profile, we would need to compute and analyze all the (potentially combinatory) Jacobians. However, we can show that all these Jacobians are the same and therefore, even in the degenerate case, we can safely study the Jacobian by using a $\mathbf{g}_q(\mathbf{x}_i, \mathbf{x}_{-i})$ as generated by Algorithm 2 except, if there are multiple best responses, Step 7–8 assign "1" only to one, randomly chosen, best response.

Theorem 13 *Given*

- *a specific sequence $q \in Q_i$ such that $x_i(q,t) = 0$,*
- *a sequence–form strategy $\mathbf{x}_{-i}(t)$,*
- *a sequence $q' \subseteq q$,*
- *the number of sequences q'' such that $q'' \cap q = q'$ and $q'' = q'|a|\ldots : a \in \rho(h), q \not\to h$ and that are best responses to $\mathbf{x}_{-i}(t)$ is larger than one,*

the eigenvalues of the Jacobian are independent from which sequence q'' is choosen as best–response.

Conclusions and future works

We developed efficient evolutionary game theory techniques to deal with extensive–form games. We designed the first replicator dynamics applicable with the sequence form of an extensive–form game, allowing an exponential reduction of time and space w.r.t. the standard (normal–form) replicator dynamics. Our replicator dynamics is realization equivalent w.r.t. the standard one. We show the equivalence for both the discrete and continuous time cases. Finally, we discuss how standard tools from dynamical systems for the study of the stability of strategies can be adopted in our case.

In future, we intend to explore the following problems: extending the results on multi–agent learning when sequence form is adopted taking into account also Nash refinements for extensive–form games (we recall, while this is possible with sequence form, it is not with the normal form); extending our results to other forms of dynamics, e.g., best response dynamics, imitation dynamics, smoothed best replies, the Brown–von Neumann–Nash dynamics; comparing the expressivity and the effectiveness of replicator dynamics when applied to the three representation forms.

Acknowledgments

Nicola Gatti is supported by MIUR under grant N. 2009BZM837 (PRIN PeopleNet).

References

Arrowsmith, D., and Place, C. 1992. *Dynamical Systems*. London: Chapman & Hall.

Cai, K.; Niu, J.; and Parsons, S. 2007. Using evolutionary game–theory to analyse the performance of trading strategies in a continuous double auction market. In *ALAMAS*, 44–59.

Cressman, R. 2003. *Evolutionary dynamics and extensive form games*. Cambridge, Mass.: MIT Press.

Fudenberg, D., and Tirole, J. 1991. *Game Theory*.

Gatti, N., and Iuliano, C. 2011. Computing an extensive–form perfect equilibrium in two–player games. In *AAAI*, 669–674.

Gatti, N.; Panozzo, F.; and Restelli, M. 2013. Efficient evolutionary dynamics with extensive–form games: extended version. http://arxiv.org/abs/1304.1456.

Kiekintveld, C.; Marecki, J.; and Tambe, M. 2011. Approximation methods for infinite bayesian stackelberg games: modeling distributional payoff uncertainty. In *AAMAS*, 1005–1012.

Koller, D.; Megiddo, N.; and von Stengel, B. 1996. Efficient computation of equilibria for extensive two–person games. *GAME ECON BEHAV* 14(2):220–246.

Kuhn, H. 1950. Extensive games. *Proc. Natl. Acad. Sci. U. S. A.* 36:570–576.

Lemke, C. E., and Howson, J. J. T. 1964. Equilibrium points of bimatrix games. *SIAM J APPL MATH* 12(2):413–423.

Lemke, C. 1978. Some pivot schemes for the linear complementarity problem. *Mathematical Programming Study* 7:15–35.

Miltersen, P., and Sørensen, T. 2010. Computing a quasi–perfect equilibrium of a two–player game. *ECON THEOR* 42(1):175–192.

Panait, L.; Tuyls, K.; and Luke, S. 2008. Theoretical advantages of lenient learners: An evolutionary game theoretic perspective. *J MACH LEARN RES* 9:423–457.

Selten, R. 1975. Reexamination of the perfectness concept for equilibrium points in extensive games. *INT J GAME THEORY* 4(1):25–55.

Shoham, Y., and Leyton-Brown, K. 2008. *Multiagent Systems: Algorithmic, Game Theoretic and Logical Foundations*.

Tuyls, K., and Parsons, S. 2007. What evolutionary game theory tells us about multiagent learning. *ARTIF INTELL* 171(7):406–416.

Tuyls, K.; Hoen, P. J.; and Vanschoenwinkel, B. 2006. An evolutionary dynamical analysis of multi–agent learning in iterated games. *AUTON AGENT MULTI–AG* 12(1):115–153.

Vermeulen, D., and Jansen, M. 1998. The reduced form of a game. *EUR J OPER RES* 106(1):204–211.

von Neumann, J., and Morgenstern, O. 1944. *Theory of games and economic behavior*. Princeton University Press.

von Stengel, B. 1996. Efficient computation of behavior strategies. *GAME ECON BEHAV* 14(2):220–246.