

Lee-Penagos, Alejandro

**Working Paper**

## Learning to coordinate: Co-evolution and correlated equilibrium

CeDEx Discussion Paper Series, No. 2016-11

**Provided in Cooperation with:**

The University of Nottingham, Centre for Decision Research and Experimental Economics (CeDEx)

*Suggested Citation:* Lee-Penagos, Alejandro (2016) : Learning to coordinate: Co-evolution and correlated equilibrium, CeDEx Discussion Paper Series, No. 2016-11, The University of Nottingham, Centre for Decision Research and Experimental Economics (CeDEx), Nottingham

This Version is available at:

<https://hdl.handle.net/10419/163012>

**Standard-Nutzungsbedingungen:**

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

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

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

**Terms of use:**

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

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

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



CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS



The University of  
**Nottingham**

UNITED KINGDOM • CHINA • MALAYSIA

Discussion Paper No. 2016-11

Alejandro Lee-Penagos  
August 2016

**Learning to Coordinate:  
Co-Evolution and  
Correlated Equilibrium**

CeDEx Discussion Paper Series  
ISSN 1749 - 3293



CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

The Centre for Decision Research and Experimental Economics was founded in 2000, and is based in the School of Economics at the University of Nottingham.

The focus for the Centre is research into individual and strategic decision-making using a combination of theoretical and experimental methods. On the theory side, members of the Centre investigate individual choice under uncertainty, cooperative and non-cooperative game theory, as well as theories of psychology, bounded rationality and evolutionary game theory. Members of the Centre have applied experimental methods in the fields of public economics, individual choice under risk and uncertainty, strategic interaction, and the performance of auctions, markets and other economic institutions. Much of the Centre's research involves collaborative projects with researchers from other departments in the UK and overseas.

Please visit <http://www.nottingham.ac.uk/cedex> for more information about the Centre or contact

Suzanne Robey  
Centre for Decision Research and Experimental Economics  
School of Economics  
University of Nottingham  
University Park  
Nottingham  
NG7 2RD  
Tel: +44 (0)115 95 14763  
Fax: +44 (0) 115 95 14159  
[suzanne.robey@nottingham.ac.uk](mailto:suzanne.robey@nottingham.ac.uk)

The full list of CeDEX Discussion Papers is available at

<http://www.nottingham.ac.uk/cedex/publications/discussion-papers/index.aspx>

LEE-PENAGOS, ALEJANDRO<sup>†</sup><sup>‡</sup>

# LEARNING TO COORDINATE: CO-EVOLUTION AND CORRELATED EQUILIBRIUM

## ABSTRACT

In a coordination game such as the Battle of the Sexes, agents can condition their plays on external signals that can, in theory, lead to a Correlated Equilibrium that can improve the overall payoffs of the agents. Here we explore whether boundedly rational, adaptive agents can learn to coordinate in such an environment. We find that such agents are able to coordinate, often in complex ways, even without an external signal. Furthermore, when a signal is present, Correlated Equilibrium are rare. Thus, even in a world of simple learning agents, coordination behavior can take on some surprising forms.

**KEY WORDS:** Battle of the Sexes, Correlated Equilibrium, Evolutionary Game Theory, Learning Algorithms, Coordination Games, Adaptive Agents.

**JEL CLASSIFICATION:** C63, C73, C72, D83

---

<sup>†</sup> School of Economics, University of Nottingham, Ph.D candidate. Email: [alejandro.lee@nottingham.ac.uk](mailto:alejandro.lee@nottingham.ac.uk)

<sup>‡</sup> Acknowledgements: Special thanks to John H. Miller for insightful discussions and comments, as well as for sharing some of his own computational routines. To Chris Starmer and Alex Possajennikov for comments on this and previous versions of the paper. Also, to participants of the Graduate Workshop in Computational Social Sciences (2015) at the Santa Fe Institute, where this project started. Financial support from the ESRC funded Network for Integrated Behavioural Sciences (NIBS) is acknowledged (ES/K002201/1).

## 1 INTRODUCTION

*“If there is intelligent life on other planets, in a majority of them, they would have discovered correlated equilibrium before Nash equilibrium”*

*-Roger Myerson, winner of the Nobel Memorial Prize in Economic Sciences<sup>1</sup>*

Aumann (1974) introduced the concept of Correlated Equilibrium (CE), which is a generalization of the traditional Nash Equilibrium (NE). Under a mixed strategy interpretation of Nash, players randomize their strategies independently of each other. In a Correlated Equilibrium such independence is not necessary: players have probability distributions based on an exogenous signal or randomization device whose distribution is common knowledge. Players map their decisions from the outcomes of such a signal to their potential actions, making their actions correlated with each other. Mutual best responses to the belief that the other players will condition their actions based on the signal is considered a correlated equilibrium.

Notice that the signal (or exogenous randomization device) has no direct influence on the payoff matrix of the game, but it can nonetheless affect the equilibrium payoffs of the players. This is not possible under NE. The definition of CE allows solutions where the signal can both affect or not the behaviour of agents. This makes it a more general concept that also includes NE, where the signals can play no role whatsoever. Perhaps this is why Myerson believes that aliens would have probably learned first to play the CE<sup>2</sup>. However, the presence of the external signal and its effect on equilibrium convergence is puzzling. It requires players to be endowed with incredible computational powers and to know the other players' payoffs. Players also have to know the signal's distribution and a specific mapping from signal to actions in order to interpret it as a recommendation of what to play. From a normative point of view, such assumptions might be adequate. But from a positive or descriptive one, it is not clear how (or if) players could actually learn this information under less straining rationality assumptions.

This paper's objective is to explore, under a canonical coordination game (Battle of the Sexes), the effects of an exogenous signal on equilibrium selection when perfect rationality assumptions are relaxed. It focuses on the behaviour of learning, adaptive, boundedly-rational agents, with an emphasis on understanding how they use the signal in order to coordinate. It takes Myerson's idea about the discovery of CE to be easier than NE as a hypothesis to be tested. Can boundedly-rational agents learn to use exogenous signals to coordinate? If so, how could this happen? Will such agents learn to condition

---

<sup>1</sup> Leyton-Brown and Shoham (2008) (p. 24) or Solan and Vohra (2002) (p. 92). Interestingly, this famous quote is often attributed to Myerson, but we couldn't find the direct source.

<sup>2</sup>In his quote, the “discovery” of the correlated equilibrium by the extra-terrestrial “players” is interpreted as them playing it in real life (i.e. to condition their actions on the exogenous signal), versus having their game theorists understand and describe the concept.

their behaviour on the signal as implied under a CE solution, or do they converge to a different equilibrium? These are the key questions explored here.

To tackle this objective, we develop a computational model with artificial adaptive agents playing a repeated Battle of the Sexes game. Analyses are made via Monte Carlo simulations. The model represents each agent as a strategy that observes inputs from the environment (such as a rival's action or an exogenous signal) and based on those observations, the agent outputs as an action in the game. We use 'finite automata', which is a mathematical model of discrete inputs and outputs that can represent boundedly rational behaviour. Such agents are allowed to adapt and change their behaviour via a learning algorithm (a 'Genetic Algorithm'). The latter simulates social learning at the population level by implementing selection and mutation processes that tend to reinforce better performing strategies and to eliminate poorer performing ones. This constitutes an evolutionary approach that explores what types of strategies emerge in the long-run.

In order to explore the impact of the exogenous signal in coordinating behaviour, computational experiments are conducted for two treatments: a baseline *No-Signal* model of the traditional game (without signal), along with a main *Signal* treatment. In the latter, agents are allowed to observe and potentially use an exogenous randomization device to coordinate.

This methodology presents several advantages for answering the above questions. First, given the interest of modelling bounded rationality, finite automata allow the representation of agents with limited memory and processing power. While they can observe the behaviour of the other agents they interact with as well as the exogenous signal, they don't have access to others' payoffs or the distribution of the signal. Hence they can only react to the observed inputs from the environment without assuming a priori complete information or infinite computational capabilities. Second, the learning algorithm implements a computational evolutionary process that allows strategies to evolve endogenously; the adaptive behaviour of the agent is given by the evolutionary dynamics of the model. This allows a wide range of strategies to potentially arise, with emerging behaviour that can potentially be difficult to predict beforehand. Such an algorithm can find strategies that were not directly specified by the researcher.

This paper contributes to the literature in its exploration of exogenous signals and correlated equilibrium by using adaptive agents. It studies the long-run effects of an exogenous randomization device on coordinating behaviour. Previous literature has also investigated coordination games by using adaptive agents, but this is the first one to allow the implementation of an exogenous signal and the exploration of its implications on equilibrium selection and evolution of individual strategies.

The model has the structure of an evolutionary tournament including two populations. In each time step, all agents in one population play a repeated Battle of the Sexes game against every other agent in the rival population. Overall scores are kept, and based on those, agents with better payoffs have a higher probability to replicate themselves and replace other agents in their own population. They undergo random mutations at the end of each time step, and the process is repeated for several thousands times simulating long term evolutionary processes.

Our results show that under both implemented treatments (with and without the exogenous signal) the system switches constantly between three

different types of equilibrium or attractors, and contrary to what was expected a priori, it never stabilises on one of them. This type of behaviour is sometimes known as ‘punctuated equilibria’, where the system remains in equilibrium for long periods of time but then presents sudden transitions into a different equilibrium. These three equilibria are i) constantly coordinating in one of the pure Nash solutions of the game, ii) symmetric alternation between the two pure Nash solutions (i.e. taking turns between the two coordination points of the game) and iii) *biased* alternation, where agents also take turns between the two coordination points, but one of them is played more often than the other. To the best of our knowledge, this is the first time that this latter behaviour has been documented in coordination experiments, whether computational or in the lab. Unexpectedly, we found no treatment differences in terms of payoffs and efficiency: both with and without the signal agents learn to coordinate quite well.

A key finding is that agents can indeed learn to condition their actions by consistently following the exogenous signal. However, even if such behaviour can be learned, the probability of it happening is very low (around 5%). While agents sometimes condition their actions based on the signal, they can also learn to alternate and coordinate their behaviour by completely ignoring it.

Hence, consistent with recent experimental literature (discussed below), our results cast doubt about CE being an accurate description of common coordination behaviour. If our adaptive computational agents can be somehow analogous to intelligent life from another planet, they will not learn CE before NE.

Finally, our methodology allowed us to identify interesting behaviour that we couldn’t predict a priori. Not only do some strategies learn to use the signal while others can coordinate by completely ignoring it, but the *same* strategy can ignore the signal, use it partially, or interpret it in different ways depending on the history of the game.

## 2 BATTLE OF THE SEXES (BOS) GAME

Figure 1 shows the payoff matrix for the traditional Battle of the Sexes (BOS) game. This game has two pure Nash strategy equilibria, with both players playing *A* (action profile (A,A)) or both playing *B* (action profile (B,B)), corresponding to the upper-left and down-right corners of the matrix respectively. In either case, one player’s expected payoff is 2 and the other’s is 3. Include now the simplest possible randomization device: both players observe the same outcome of a fair coin toss before deciding their actions, with a 50% probability of observing H (*Heads*) and 50% T (*Tails*).

		Column Player	
		A	B
Row Player	A	2,3	0,0
	B	0,0	3,2

Figure 1: Payoff Matrix in Battle of the Sexes Game

Traditionally, H or T is interpreted as an exogenous signal or a non-binding recommendation for players on what actions to choose. For example, with probability 0.5 both players are ‘recommended’ to play A (i.e. the recommended action pair is (A,A)) when, say, Heads shows up and (B,B) otherwise (when Tails). This is a *correlated strategy*, which is given by this joint distribution over the set of pure strategy pairs. Notice that in this case, the expected payoff for both players is 2.5 (since each outcome *AA* or *BB* would be played with 50% probability), which differs from the expected payoffs of any of the two NE<sup>3</sup>.

This correlated strategy is also a CE because no player wishes to depart from following the recommendation. For example, when Heads shows up with recommendation (A,A) and given that player Column will follow it, player Row would decrease its payoffs by not playing what is recommended: if Row decides to play B, his payoffs would be zero instead of two. The same is true for player Column, whose payoffs would go from three to zero in the analogous situation.

The CE concept requires each player to assume that the rival will follow the recommendation given. It also requires common knowledge of the distribution of signal as well as every other agents’ payoff. Here we will relax these assumptions. As explained in section 4.2, in our model the signal will be observed by the agents without any common knowledge assumption, and it is the dynamics of the model that will determine if they learn to use it consistently to coordinate or not. Also, there will not be any given function mapping the signal to particular actions (i.e. no recommendations): whether agents learn to give particular meanings to the signal or not will be determined endogenously by the evolutionary process of the model.

The payoffs of the game can be formalized graphically as in Figure 2. The line  $\overline{ABC}$  is the boundary of the convex hull, so all payoffs combinations on the line or inside of the triangle are feasible with appropriate randomization. The maximum attainable payoff for a single player must occur at one of the vertices of the convex hull (i.e. when a pair of pure strategies is played). In this case those points are  $A = (2,3)$  and  $B = (3,2)$ . In this BOS game, A and B are also the two pure Nash equilibria. The line  $\overline{AB}$  forms the set of Pareto optimal solutions. Point  $D = (2.5, 2.5)$  is the CE discussed earlier. An interesting characteristic of this point is that it is not only Pareto efficient, but is also an *egalitarian* equilibrium: a priori, before the coin toss, both players have the same expected payoffs<sup>4</sup>.

To avoid confusion in the analysis that follows, we need to carefully specify what we mean by a CE in our model, or by ‘behavior consistent with CE’. Technically, many solution concepts, including pure Nash, are also a CE.

---

<sup>3</sup> Although this paper will not allow the possibility of mixed strategies, is worth noting that these expected payoffs cannot be obtained by players randomizing on their own (i.e. without the signal). The coin toss in this case allows payoffs that cannot be obtained under a mixed Nash equilibrium concept.

<sup>4</sup> Aumann (1987) suggests the fair coin toss as one of the most simple randomization devices, making it a good candidate for studying the emergence of CE. Another equilibrium studied in the literature is for the Chicken game (Duffy and Feltovich, 2010), with the characteristic that the CE is outside of the convex hull of NE (i.e. the randomization allows higher Pareto efficient payoffs). However, as those authors argue, such CE can be more difficult to learn (at least for humans) because it requires three recommendation profiles, instead of the two implemented here. Cason and Sharma (2007) find that humans’ difficulty in learning a CE comes from the uncertainty about their rival’s actions, not a lack of incentives (i.e. higher payoffs). One objective here is to test a CE that could arguably be the easiest to learn.



However, the interest here is to focus on CE that requires agents conditioning their actions on the signal. So for a CE, we will require agents to condition on the fair coin toss without ignoring it. At the aggregate level this implies payoffs close to the egalitarian equilibrium; at the individual level, as in the correlated strategy above, using the signal as if a recommendation profile is being followed.

Other behavior, even if under traditional theoretical assumptions (which are relaxed in our model) could also be labeled as CE, will be referred to independently in order to maintain focus on this particular form of signal use.

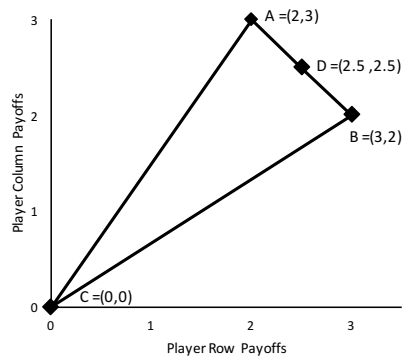


Figure 2: Set of attainable payoffs of BOS game under a correlated strategy pair.  $\overline{ABC}$  is the boundary of the convex hull, hence any payoffs on or inside of this hull are attainable with the appropriate randomization. Point  $D$  represents the correlated equilibrium given by a fair coin toss as the randomization device.

For a more formal presentation of the one shot game and some equilibrium concepts, see Appendix 7.1.

### 3 RELATED LITERATURE

#### 3.1 On Correlated Equilibrium and learning

Aumann (1974) introduced the concept of CE into the literature and refined it in Aumann (1987), showing that Bayesian rationality implies convergence to a CE. However, this required players to have the same prior beliefs regarding the distribution of the exogenous signal. Some following papers focused on giving conditions or learning rules for achieving convergence. In Foster and Vohra (1997), such convergence is based on players making “calibrated forecasts”. This implies evaluating the complete past actions of all rivals, and using this to make perfect probabilistic forecasts that match beliefs with randomized strategies. Fudenberg and Levine (1999) presented an alternative mechanism requiring similar memory capabilities. Hart and Mas-Colell (2000) introduced convergence via “regret”, with players making *better* choices instead of using *best* responses (i.e. they switch to actions that would have given higher payoffs than the ones used in the past). This latter approach relaxes some of the rationality assumptions in previous work, but still requires

players to have a complete memory of all past actions and calculate the potential payoffs of all of the strategies that could have been played under all potential scenarios. These approaches require very sophisticated players, with unbounded memory and computational capabilities, playing indefinitely. In contrast, the approach here is to model agents with limited memory and no prior beliefs, and test whether evolutionary learning processes at the population level can lead them to learn the CE.

Recent experiments in the lab have focused on CE. These studies have been conducted by Cason and Sharma (2007), Duffy and Feltovich (2010), Bone et al. (2013), Duffy et al. (2014) and Anbarci et al. (2015). A key result arising in all of them is that while some subjects do follow the recommendations given, they do so inconsistently, casting some doubt on the descriptive power of the CE concept<sup>5</sup>. However, as conjectured by Cason and Sharma (2007) in their conclusions, perhaps in longer time spans subjects might learn to consistently follow the recommendations. The evolutionary approach with adaptive agents presented here addresses this issue by conducting long-run analyses that would be impossible to conduct in the lab. Also, it is worth noting that all the experiments above give subjects common knowledge about the distribution of the recommendations as well as what is being recommended to the rival. While such information is useful in helping subjects understand the experiment, how is it that agents come to know such information in a different environment?

### **3.2 On methodology**

This paper uses artificial adaptive agents to study the learning and evolution of behavior consistent with CE. Modelling artificial adaptive agents serves as a great compliment to theoretical analysis in economic theory (Holland and Miller, 1991), and it has been used in a wide range of social science topics like market institutions (Gode and Sunder, 1993), pricing (Arifovic, 1994), auctions (Andreoni and Miller, 1995), the evolution of norms (Axelrod, 1986), elections (Kollman et al. (1992)), political institutions (Kollman et al., 1997), loyalty in fish markets (Kirman and Vriend, 2000) and the emergence of communication (Miller et al., 2002), among many others.

Agents presented in this work are boundedly rational with limited information and memory. They are embedded with a mechanism that promotes constant adaptation to their changing environment. Since all agents adapt to each other at the same time, they constitute a co-evolving complex adaptive system. Such adaptive behavior is modelled by means of a genetic algorithm (Holland, 1992), which captures the idea of social learning: strategies that are successful are more likely to be copied by other agents and hence spread in the population, but strategies that are unsuccessful are more likely to be distorted in the learning process. The algorithm strikes a balance between exploration and exploitation (i.e. looking for new solutions versus

---

<sup>5</sup> Our evolutionary methodology makes it impossible to make quantitative comparisons with the results obtained in the short time span possible in the lab. However, in section 5, we observe qualitative patterns that also emerge in these experiments, giving some external validity to the model presented. The experimental literature is also relevant because results in the lab can inspire new scenarios to explore computationally and vice versa. It is our belief that complementarities and mutual feedbacks exists in social sciences between studies conducted with humans and machines (Duffy (2006) or Poteete et al. (2010) present overviews of this methodological complementarity. Andreoni and Miller (1995) is an example of lab experiments working in tandem with computational simulations).

exploiting the ones that have already been found), which constitutes a classic conundrum in problem-solving (Holland, 1992; Holland et al., 1986).

Each agent is defined as a finite state automaton. Rubinstein (1986) was the first to introduce automata into game theory as representations of strategies. Miller (1988) introduced the idea of using evolutionary algorithms to model adaptive learning in games (Miller (1996), Ioannou (2013) and Zhang (2015)). Some recent studies have explored the use of automata in coordination games (such as Browning and Colman, 2004; Hanaki, 2006; Ioannou and Romero, 2014a; Ioannou and Romero, 2014b) but no one has studied exogenous signals or CE.

Here we explore with adaptive agents the long run emergence of CE behavior. Arifovic et al. (2015) used *individual* learning to see if adaptive agents can replicate quantitatively the short-term behavior of subjects in the laboratory, including exogenous recommendations. In contrast the approach here uses *social* learning at the population level to focus on the long-term evolution of signal conditioning.

## 4 THE COMPUTATIONAL MODEL

### 4.1 Overall structure

The game used in this paper is the repeated Battle of the Sexes (BOS) as presented in section 2. Each agent represents a strategy, and agents face each other in a computational tournament.

More specifically, agents are represented as finite automata (their formalization explained in detail in section 4.2). The model has two populations, *COL* and *ROW*, each one consisting of  $N$  agents. Each time step of the model is called a generation, denoted as  $t$ . At each  $t$ , each agent in population *COL* plays  $R$  rounds of the BOS game against each other agent in population *ROW*. The average score (payoffs) of each agent is recorded across all  $R \times N$  rounds of play in one generation. Agents select their strategies by imitating the strategies used by other successful agents, with the average score being the (fitness) measure used of success. Hence, strategies with lower scores will tend to disappear from the population while those with higher scores will tend to spread. This is due to the learning algorithm (detailed in section 4.3) giving successful strategies higher probability of being copied by other agents. This learning happens at the end of each generation, with agents copying only strategies that are in their own population, thus the *ROW* and *COL* populations evolve independently of each other.<sup>6</sup>

The computational experiments conducted here consist of two main treatments: *No-Signal* and *Signal*. Under *No-Signal*, agents play without any randomization device or exogenous signal. In the main treatment, *Signal*, agents play under the same game structure, but are allowed to observe an exogenous signal (given by the fair coin toss) at the beginning of each round.

---

<sup>6</sup> The choice of the structure of the game, mainly repeated interactions (instead of one-shot) and having two populations instead of one, makes learning potentially easier and should give the emergence of the CE the best possible chance. Experiments conducted by Duffy and Feltovich (2010) show that humans in the lab learn more frequently to follow the exogenous signals in coordination games when they play repeatedly versus playing in one-shot interactions.

## 4.2 Artificial agents as finite automata

Each agent is defined as a class of finite automata using a Moore machine (Moore (1956)), which is a mathematical model with discrete inputs and outputs<sup>7</sup>. The system can be in any of a finite number of internal configurations, called “states”. States summarize the past set of inputs and determine the automaton’s behavior for subsequent outputs.

A finite automaton can be described as a four-tuple  $(Q, q_0, f, \tau)$ , where

- $Q$  is a finite set of internal states,
- $q_0 \in Q$  is specified to be the initial state,
- $f: Q \rightarrow A_i \in \{A, B\}$  is an output function that maps each state into an action of the machine, and
- $\tau: Q \times W \rightarrow Q$  is a transition function assigning a state to every two-tuple of state and observed input.

Here,  $W = A_{-i} \in \{A, B\}$ , where  $A_{-i}$  is the action implemented by the other agent. In this case the only input used by an agent to decide its next action is the action implemented by its rival. In the BOS such input can be A or B, giving agents two potential inputs to respond to. This is how the agents are implemented for the *No-Signal* treatment.

In the *Signal* treatment, each automaton is allowed to respond to four different inputs. Let  $S \in \{H, T\}$  be an exogenous random signal with a probability distribution  $\left[\frac{1}{2}, \frac{1}{2}\right]$  (e.g., a fair coin toss showing either Heads (H) or Tails (T)), and having the same value H or T for any pair of interacting agents at a given round (i.e. both agents observe the same signal). Thus, in this treatment  $W = A_{-i} \times S$  with  $W \in \{(A, H), (A, T), (B, H), (B, T)\}$  giving all four possible combinations of the other agent’s action and observed signal.

An intuitive way to describe an automaton is by using a transition diagram. Figure 3 shows two examples of such diagrams. The nodes in the transition diagrams represent the internal states. The arrows originating from each node represent the transition function with the labels showing the input (rival’s action and signal) required for a transition. The arrows point towards the state that the automaton transitions to after observing the corresponding input. The initial state of the machine is given by the “start” arrow.

The automaton in Figure 3 (a) for the *No-Signal* treatment shows a strategy that starts by playing A in the first round. Afterwards, it does the same as the rival did in the last round: whenever it observes A it transitions to the state playing A, and whenever it observes B it transitions to the state playing B. This is the famous Tit-for-Tat strategy (Axelrod (1980)). In the *Signal* treatment (Figure 3 (b)), transitions are coded using two letters, the first representing the rival’s last action (A or B) and the second representing the observed signal (H or T). So a transition showing, say, AT, means that such a transition occurs when the agent observed the rival playing A in the last round and the signal is T for the current one. There are four possible transitions for each node in the diagram. The strategy here starts playing A and when it observes a signal of T, regardless of the rival’s past action or the machine’s current internal state, it

---

<sup>7</sup> There are other types of finite automata such as Mealy machines. Choosing Moore machines as the type of automata implemented is due to it being the standard in previous game theoretical literature. We see no evident reason to deviate from this convention.

will play A. This is easily noticed by observing that all the arrows that have a signal of T go into the initial state. Similarly, whenever it observes H, regardless of the rival's action, it plays B. This strategy gives a consistent interpretation of the signal: play A when T, B when H. This is one possible strategy that could be consistent with CE behaviour.

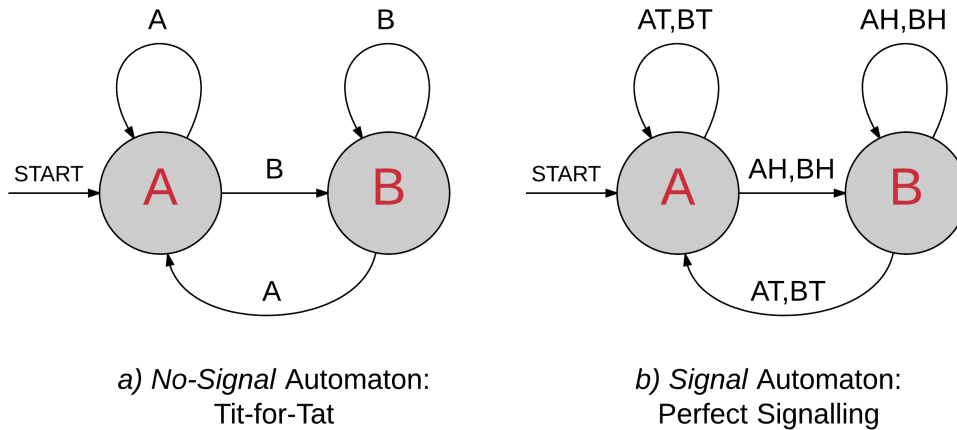


Figure 3: Examples of automata for both No-Signal and Signal treatments.

In order to use the learning routines (explained in section 4.3) the automata need to be coded as finite length strings. Figure 4 shows the coding for both treatments. The *No-Signal* automata is coded as a 25-length string, where the first element provides the initial state of the machine (Figure 4(a)). Then, there are eight three-element packets, each representing one of the eight internal states of the automaton<sup>8</sup>. In these packets, the first element gives the action the agent takes when it is in that particular internal state (i.e. to play either A or B). The other elements are the transitions to make when observing the different inputs (i.e. the rival's action): the second element is the transition when the rival is observed to play A, and the third element is the transition when observed to play B (Figure 4(b)). The coding for the *Signal* treatment is very similar, with the difference that it requires a longer string (41 elements instead of 25). This is because including the signal allows four possible inputs, requiring four transition per internal state (instead of two). Hence for each state, as in Figure 4(d), the first element is the action to be taken, and the following elements are the transitions for all four possible combinations of the rivals' action in the last round and the observed signal in the current.

---

<sup>8</sup> The number of states used in the machines is in line with previous literature. For example, Ioannou (2013) also uses eight internal states arguing that it allows for a variety of automata that can incorporate a diverse array of characteristics. It is worth noting that more complex machines (more states) do not necessarily mean better strategies. As pointed by Rubinstein (1986), more complex plans of actions are more likely to break down, are more difficult to learn, and can require more time to be executed. Gigerenzer et al. (2011) has several examples of simple rules of thumb that perform better than complex strategies.

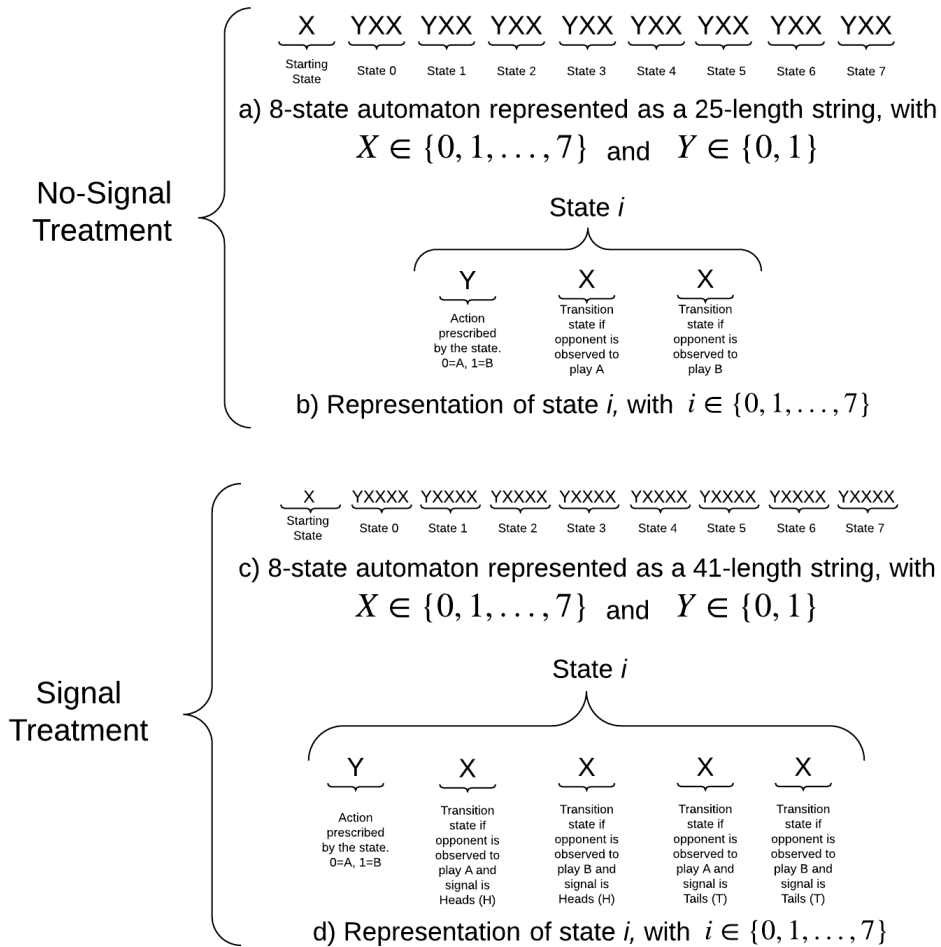


Figure 4: Coding of automata for both No-Signal and Signal treatments.

There are some important technical points inherent to the use of automata. Notice that the machines don't have any sort of "expectations" of what the rival will do and that their behavior is purely backwards looking, which is one way to represent simple, boundedly-rational strategies in evolutionary processes. Also, although no separate computational memory is implemented, the internal state of the machine contains the relevant history of the game. A strategy that is based on the past  $n$  moves of its opponent will require a maximum of  $2^n$  internal states: for example, the Tit-for-Tat strategy requires the automata to remember only the last action of the opponent, hence it requires two states. Even if the automata here is modelled with eight internal states, only a subset of these states may be accessible to a machine given the starting state and transitions. The number of potential configurations of machines is rather large. In the *No-Signal* treatment, there are  $8^{16} \times 2^8$  different arrangements of the strings. However, since many of the configurations lead to the same behavior, the number of unique strategies is lower. For example, two-state machines have  $2^7 = 128$  possible arrangements (genotypes) but only 26 unique strategies (phenotypes)<sup>9</sup>. Finally and related to

<sup>9</sup> For 3-state machines (with also two inputs and two outputs) the number of unique strategies is 5,832. Notice the exponential growth in the number of possible phenotypes. This

the latter, automaton theory, as in Harrison (1965), proves that isomorphic automata that represent the same behavior can be mapped to a minimal state machine in the canonical form. This means that many different machines can lead to the same behavior, all of them being able to be represented by a single ‘minimal’ automaton. These are referred to as “behaviorally equivalent” or “minimized” machines.

### 4.3 Evolution of strategies

#### 4.3.1 Motivation for the learning mechanism

The learning algorithm used in this paper is derived from a class of optimization routines from computer science called genetic algorithms (GA), introduced by Holland (1975). GAs are computer programs that mimic the processes of biological evolution in order to solve problems and to model evolutionary systems. We use GAs for two main reasons: its technical advantages and its analogy as a learning mechanism reflecting bounded rationality.

The algorithm has several advantages over other optimization methods. It is designed to work well in *difficult* domains, meaning domains that involve discontinuities, nonlinearities (many local optima), noise and high dimensionality (these issues arise in the strategy space in the tournament analyzed here). Contrary to calculus-based methods that require derivatives in order to perform an effective search for better structures, GAs require payoffs associated with the individual strings, making it ideal for game theoretical environments with their well-defined payoffs structure. All of the above makes GAs a more canonical optimization method than many other search schemes<sup>10</sup>.

Evolutionary processes such as a GA explicitly model a dynamic process describing how agents adjust their choices over time by learning from experience; this makes the GA a useful tool for observing the learning (or lack thereof) of coordinating behavior with an exogenous signal. In the same line as Kandori et al. (1993), this evolutionary approach gives a concretely defined, step by step process of how an equilibrium can emerge based on trial and error mechanics. Even if biological interpretations are usually given to such processes, the algorithm’s processes can be reinterpreted as bounded rationality, reflecting the limited ability on the player’s part to receive, decode and act upon information they get in the course of the game. As in Kandori et al. (1993) three main hypotheses are relevant and related to this learning interpretation, reflecting its adequacy in order to model adaptive, boundedly rational agents. First, the *inertia* hypothesis holds since not all players react instantaneously to their environment. This is because given the imperfect observations agents have (for example, regarding payoffs and strategic choices of other agents), changing one’s strategy can be costly. Second, the *myopia* hypothesis holds since there is substantial inertia in the system with only a small fraction of agents changing their strategies simultaneously, resulting in agents making only moderate changes. The myopia hypothesis also captures a key factor in *social* learning: imitation or emulation. Agents learn what are

---

makes calculations on the exact number of possible strategies for machines with more internal states increasingly costly in computational terms.

<sup>10</sup> For further discussions on genetic algorithms, see Mitchell (1998).

good strategies in a complex environment (where they cannot calculate best responses) by observing what works well for others. In such an environment strategies that remain effective in the present are likely to remain effective in the near future. Also, myopic agents do not take into account the long-run implications of their actions or strategies. Finally, the *mutation* hypothesis holds given that with some small probability agents will play an arbitrary strategy, capturing the exploration aspect of most learning processes.

#### 4.3.2 Details of the Genetic Algorithm implementation

The mechanics of the implemented GA (for both *No-Signal* and *Signal* treatment) are as follows: two populations (ROW and COL) are randomly initialized with 40 agents each at  $t=1$  (first generation). This initialization consists of generating for each agent a random finite-length string automaton as in Figure 4 (with uniform probability across the alternatives)<sup>11</sup>. Then each automaton is tested against the environment: this consists of each agent in population ROW playing 50 rounds of the repeated BOS game against each of the 40 agents in COL population. Scores are stored for all automata, with the score for each agent being the average payoffs earned across all games.

Two new *offspring populations*, each with 40 agents, are created based on the current *parent populations* (i.e. the populations existing at the beginning of the generation). Each population evolves independently, so the offspring of the COL population will be based only on the parent COL population (the same applies for ROW). Offspring populations are created based on two operators: *selection* and *mutation*. For selection, the top 20 scorers are chosen and given a copy in the new population. The other 20 needed to keep populations constant are chosen via pairwise tournaments by randomly picking two agents (with replacement), and keeping the one with the highest score. Such tournaments are repeated 20 times in order to keep population size constant.

Before moving on to the next generation, the 20 strategies picked via the pairwise tournament go through mutation process. Each automaton has a 0.5 probability of being randomly altered. If a strategy undergoes mutation, one of the internal states is randomly selected and with a 0.5 probability the action of that state is changed (thus, if the state had an action of A, it is changed to B and vice versa); otherwise, a randomly chosen transition (from the chosen state) is changed with uniform distribution for the alternatives<sup>12</sup>.

---

<sup>11</sup> Randomly generated populations will favour minimized (behaviourally equivalent) machines that represent strategies with only one internal state (i.e. always play A or always play B). When the maximum internal states allowed is equal to two, the probability of generating a machine that always plays A is 31% (analogous for always playing B). When three internal states are allowed, this probability is 20%. Making such calculations for more internal states becomes increasingly costly; however, the dynamics of the GA will quickly start favouring strategies that perform better.

<sup>12</sup> There are other ways to implement selection and mutation. GAs are a broad class of algorithms with many variations, but fortunately they are fairly robust to different parametric and algorithmic choices. The mutation parameters and mechanism used are the same as in Miller et al. (2002) and Miller and Moser (2004). In general, within reasonable changes, results will be consistent. However, if taken to an extreme, too small mutation rates eliminate exploration and will lead the system to converge based only on the selection process. If mutation is too high, the system will always be exploring, unable to settle down and exploit information. The chosen mechanism tends to be in a reasonable “sweet spot” to balance this out.



Finally, once both ROW and COL offspring populations have been created, scores are reset to zero and a new generation of the algorithm is begun (i.e. agents are again tested against the environment, scores are assigned, and populations undergo selection and mutation). An overview of the whole process is given in Figure 5.

- 1) Initialise two random populations (ROW and COL) with 40 agents each. Set  $t=1$  (first generation)
- 2) Test each agent against the environment: play 50 rounds of BOS against each agent in the rival population, saving average scores.
- 3) For ROW population, form a new population of 40 agents in the following way:
  - a) Copy top 20 scorers from old population (will also be potential parents)
  - b) Pairwise tournament: choose randomly 2 potential parents from the population of 20 copied in (a), with replacement. The one with the highest score gets one child copy of itself
  - c) With 50% probability, mutate the child:
    - i) Randomly choose one internal state
    - ii) With 50% probability, switch the action of that state
    - iii) If didn't change action in step (ii) (50% prob.), randomly choose one transition of the state and change it with uniform probability across alternatives.
  - d) Repeat steps (b) and (c) until the new ROW population has 40 agents.
- 4) Do step (3) for COL population
- 5) Increment  $t$  by 1 (next generation), reset scores to zero and iterate (go to step (2)).

*Figure 5: Structure of the evolutionary process (works the same for both No-Signal and Signal treatment)*

## 5 RESULTS

Given the model we can analyze its behavior. The following five questions address the overarching research goals presented in the introduction, serving as a roadmap for the evidence ahead. They will be answered in the order presented.

- 1) Will the system converge to an equilibrium?

A priori, is not clear if an equilibrium will emerge. We hypothesize that without the signal agents will converge into one of the pure Nash equilibria. With it, our hypothesis is that they will converge in Turn-Taking (alternation), taking turns symmetrically in both coordination points of the game. For both treatments the hypothesis is that the system will stabilize in the corresponding equilibrium and remain there.

- 2) Will the presence of the signal allow agents to coordinate more easily? That is, will the system be more efficient when the signal is included?

We hypothesize that when the signal is included, agents will miscoordinate less often leading to higher payoffs.

- 3) Are there other treatment differences, if any, in terms of the aggregate behavior of the system?

We have no other a priori hypotheses regarding treatments differences besides the ones addressed in questions 1 and 2, but we leave the possibility for unexpected results. With the power of hindsight, we know that there are indeed other differences that are worth exploring once answers to questions 1 and 2 above are known.

- 4) Conditioned on observing Turn-Taking (alternation) as hypothesized in question 1, will agents be actually conditioning on the signal in a way consistent with CE?

This question might seem subtle, but its analysis is key to understanding the emergence of CE. Notice that agents might alternate or take turns in the two coordination points by either using the signal or by completely ignoring it. Both types of behavior would seem similar at the aggregate level, but only conditioning on the signal would be consistent with CE as defined here. We hypothesize that agents will learn to condition their actions based on the signal.

- 5) At the micro level, how are agents coordinating? That is, how do we characterize the strategies that evolve?

Analysis of questions 1 to 4 are made at the aggregate level of the system (e.g. average payoffs, coordination rates). But one of the advantages of using automata and computational methods is that we can directly observe each and every strategy in the system at any point in time. Here we use a methodology based on *pairs of interacting* strategies to characterize them and understand their exact behavior. A priori, given the immensity of the possible strategy space, we don't have any particular expectation of the type of strategies that would evolve besides the ability to invoke both pure Nash and alternating behavior. However, as we will see, novel and interesting behavior evolved that we didn't predict beforehand.

### 5.1 Regimes and epochs

We start by focusing on what type or types of equilibrium are selected under the *No-Signal* treatment. Figure 6 shows the average payoffs obtained by each population across all rounds of play. Five panels are shown, each one of them corresponding to a different run of the model. Some key patterns can be observed and some characteristics inferred based only on the average payoffs.

Note that the system never fully stabilizes. Instead, it is characterized by punctuated equilibria: the system locks for several generations in a kind of stasis where average payoffs per population are quite stable, followed by a sudden transition into a different (and similarly stable) configuration<sup>13</sup>.

---

<sup>13</sup> The assertion that the system “never” stabilises is based on longer runs. Some of the earliest literature on similar models ran simulations for around 50 generations. Recent work has

Three kinds of equilibrium behavior are identified. Remembering that both game's pure Nash equilibria have payoffs of (3,2) and (2,3), the run in the top panel of Figure 6 shows consistent coordination on either (A,A) or (B,B). In this run, one population is consistently receiving average payoffs very close to three and the other very close to two. Thus, one population is 'dominating' the other in terms of payoffs. The transitions here only change which population is getting the higher payoffs.

The second equilibrium behavior observed, for example, on the third panel around the 1,000 generations mark, has both populations obtaining average payoffs close to 2.5. Given the structure of the model, without the exogenous signal this means that the agents have found a way to coordinate on some sort of turn-taking behavior. They are alternating symmetrically between the two coordination points, although it is not clear if they are alternating each turn. They could, for example, by playing three times in a row (A,A), then three times in a row (B,B), and so on.

The third equilibrium that arises in the model was not foreseen. It can be observed in the bottom panel, around generations 1,100 to 1,700. Here agents use 'biased turn-taking': although they take turns, it is not symmetric. Agents are playing, for example, two rounds at (A,A), followed by one round of (B,B) and then back to (A,A). This gives both agents a chance to play to their preferred coordination point, but one of them having its way more often. This is the first time such behavior has been documented in a BOS game, either in simulated or experimental data. The micro analysis showing exactly what strategies emerged for all three equilibria will be done section 5.5.2, allowing us to understand how such coordination happens.

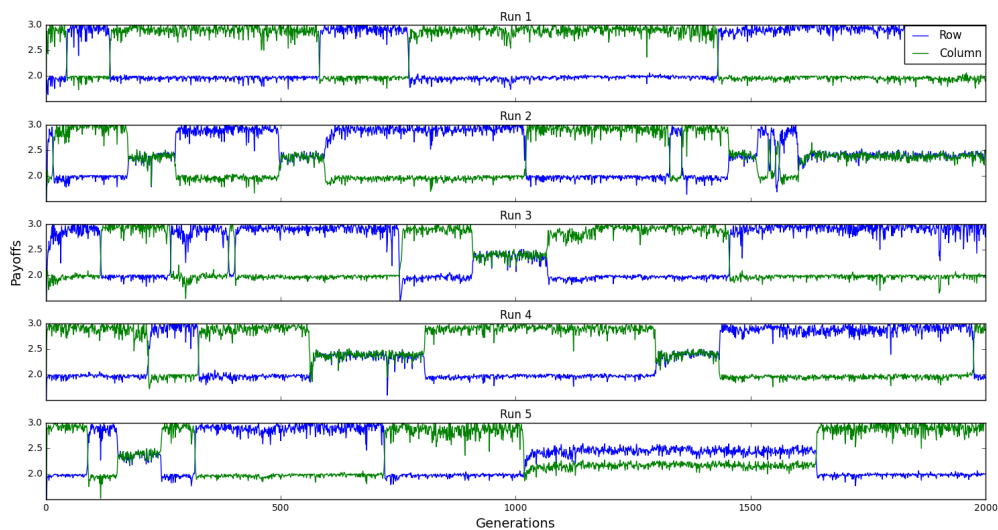


Figure 6: No-Signal treatment. Average payoffs per population. Each panel is one different run of the model, each consisting of 2,000 generations.

It is convenient to have a formal way to describe and name these equilibria: each generation,  $t$ , will be classified under one of the following regimes based

---

used between 1,000 and 2,000 generations. Besides the five simulations, the model has been run several times up to 5,000 and 10,000 generations. One very long simulation that will be reported below was run for 100,000 generations. In all runs the system displayed punctuated equilibria.

on the a-posteriori probability of observed play. Let  $AA_p^t$  be the percentage of rounds for any pair of agents playing (A,A) during generation  $t$ , and  $BB_p^t$  the analogous for (B,B). Then each generation is classified into one of four regimes according to the following rules<sup>14</sup>:

- *Domination A (B)*: if  $AA_p^t (BB_p^t) > 0.8$
- *Turn-Taking*: if  $(0.4 > AA_p^t < 0.55)$  and  $(0.4 > BB_p^t < 0.55)$
- *Biased Turn-Taking A (B)*: if  
 $(0.15 > AA_p^t (BB_p^t) < 0.4)$  and  $(0.55 > BB_p^t (AA_p^t) < 0.80)$
- *Other*: if none of the above.

An *epoch* is defined as a streak of consecutive generations under the same regime. Technically, it is a window of at least ten generations with the same regime where no more than three are being classified under a different regime (hence allowing for some “mistakes”). For example, 500 generations in a row classified under the regime “Domination A” (allowing for a few mistakes) is considered as **one** ‘Domination A’ epoch.

In order to have representative measures of the system’s behavior, one very long simulation (with  $t=100,000$ ) was run for each treatment<sup>15</sup>. Compared to the  $t=2,000$  of initial simulations, the longer time span gives us a good measure of the system’s statistical properties. All of the following data for each treatment is based on the corresponding long simulation<sup>16</sup>.

Based on such long simulations, less than 1% of generations are classified under the ‘Other’ regime. So the system under the *No-Signal* treatment can be

---

<sup>14</sup> The threshold values for each regime were chosen in order to allow a convenient classification, and the analysis is robust to reasonable changes.

<sup>15</sup> Having one very long simulation instead of aggregating several short ones for the main analysis was chosen for a reason: as will be seen below, some epochs can be rather long, characteristic that would be lost with short simulations.

<sup>16</sup> Although one might initially have concerns for the effects of the random initial conditions, given enough time and due to the switch between epochs (i.e. the phase transitions), the system will eventually forget its past. Each type of epoch (i.e. regime) can be seen as an attractor of the model, and by visiting them all the system is no longer dependent on the initial conditions. This would be different if the system would lock in one of the attractors forever, which would make initial conditions critical.

accurately described in terms of the three main regimes Domination, Turn-Taking and Biased Turn-Taking.

What is the equilibrium behavior of the system when the signal is included? Surprisingly, it is very similar to the *No-Signal* treatment. One can grasp an intuitive feeling for this by observing appendix 7.2, where figures for the 100,000 simulations and five short ones for the *Signal* treatment are presented. The reader will notice that the payoffs present very similar patterns compared to the *No-Signal* treatment. Formally, based on the corresponding 100,000 generations simulation, the system can also be classified in more than 99% of the time in one of the three main regimes, and constant transitions between them are also observed. This means that at the aggregate level, both with and without the signal the model presents similar behavior in terms of the regimes that emerge. Other treatment differences will be addressed below, including the probability of finding the system in each regime (confirming this result).

The evidence so far can be summarized as follows:

**Result 1:** *The behavior of the system can be described in terms of three main regimes: Domination, Turn-Taking and Biased Turn-Taking. The system never stabilizes in one particular regime, but instead presents transitions switching from one long epoch to another in short time spans. This applies for both Signal and No-Signal treatments<sup>17</sup>.*

## 5.2 Efficiency

The next question we consider is the efficiency of the system. Table 1 presents the average payoffs in the long run as well as the average coordination rates. The latter is measured as the percentage of rounds across all generations where any pair of agents play a coordination point (either (A,A) or (B,B)). In terms of payoffs both treatments have virtually the same value of 2.4, which is very close to the Pareto optimal of 2.5<sup>18</sup>. Coordination rates also show that the system is highly efficient. In both treatments agents play one of the pure Nash strategies (i.e. a coordination point) in more than 95% of rounds. Comparing this with the expected coordination rates for agents playing mixed strategies (48%) or even playing randomly (50%), it can be seen that the system is equally efficient with or without the use of the signal. Contrary to what was hypothesised a priori, the signal doesn't really help agents solve the coordination problem.

**Result 2:** *Under both treatments the system is quite efficient: the probability of agents coordinating in one of the two pure Nash equilibria is close to 95% with and without the signal. Payoffs are virtually the same and very close to the Pareto optimal of 2.5, so we conclude that there are no*

---

<sup>17</sup> Simulations using an alternative selection mechanism also have been run. Instead of selecting 20 top scorers to go directly into the next generation and then using a pairwise tournament, the alternative was to conduct the tournaments directly for the whole population, without guaranteeing any strategy a direct copy. Simulations are robust to this result, namely the regimes observed and the constant transitions between them.

<sup>18</sup> Average payoffs by population are, for the *Signal* treatment 2.38 and 2.43, and for the *No-Signal* treatment, 2.26 and 2.48. Due to the large amount of observations, differences are statistically significant, although they seem relatively small in economic terms. Such small differences can occur due mainly to the presence of some very long epochs, particularly for the Biased Turn-Taking regime (as shown below).

treatment effects in terms of payoffs or efficiency. Agents learn to coordinate equally well with or without the exogenous signal.

Average Payoffs		Average Coordination Rate	
<i>No-Signal</i>	<i>Signal</i>	<i>No-Signal</i>	<i>Signal</i>
2.36	2.4	95%	96%

Table 1: Average Payoffs and Coordination Rates for both No-Signal and Signal treatments. Treatment differences are barely noticeable.

### 5.3 Probabilities of each regime

We turn now to the differences in regime frequencies. Figure 7 presents the probability of randomly choosing one generation and having it classified under each regime. The percentages presented are equivalent to the ratio of the number of generations classified under each regime to the total number of generations in the run (here  $t=100,000$ ). This provides a measure of how much time the system spends in each regime. For easy of exposition, notice that ‘Domination A’ and ‘Domination B’ are aggregated simply as ‘Domination’ (the same applies to Biased Turn-Taking).

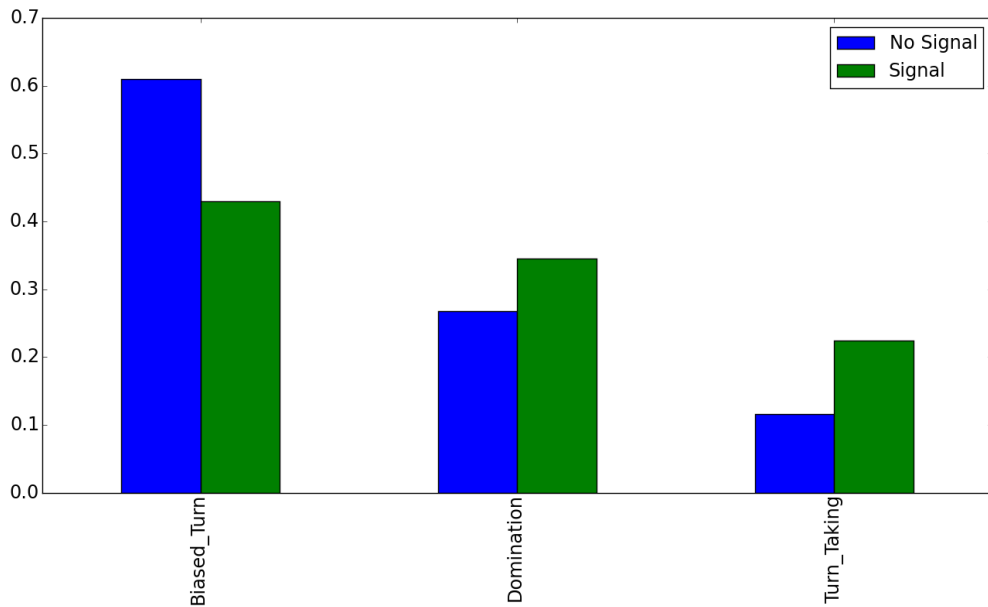


Figure 7: Percentage of the time that the system spends in each regime. Measured as the ratio of generations classified under each regime to the total generations in the run ( $t=100,000$ ).

Figure 7 shows that Turn-Taking is the least frequent of the three regimes, both with and without the signal. This suggests that CE may not be a more likely equilibrium concept than pure Nash. There’s also no evidence that Turn-Taking would be learned “first”. All simulations ran started with a short period of learning (usually no more than ten generations) followed by a Domination epoch. This is due to the random generation of automata favoring strategies that always play A or always play B (as mentioned before). So in this model,

behavior consistent with pure Nash equilibrium is both more frequent and happens before any kind of Turn-Taking. Figure 7 also shows that the system spends most of the time in a Biased Turn-Taking regime under both treatments. Why is this the case?

One potential explanation for the prevalence of Biased Turn-Taking is that the system transitions more often into these epochs than into the others. Figure 8 shows the total number of transitions the system underwent (a), and how are those distributed across the three regimes, i.e. the percentage of transitions into each regime (b). It can be seen that even if the system transitions more often under the *No-Signal* treatment, the distribution is the same under both treatments. For both treatments, Biased Turn-Taking is the regime to which the system transitions into *least* frequently. If Biased Turn-Taking is the more frequent regime, but also the one to which the system transitions into less frequently, the length of the epochs must be driving our results<sup>19</sup>.

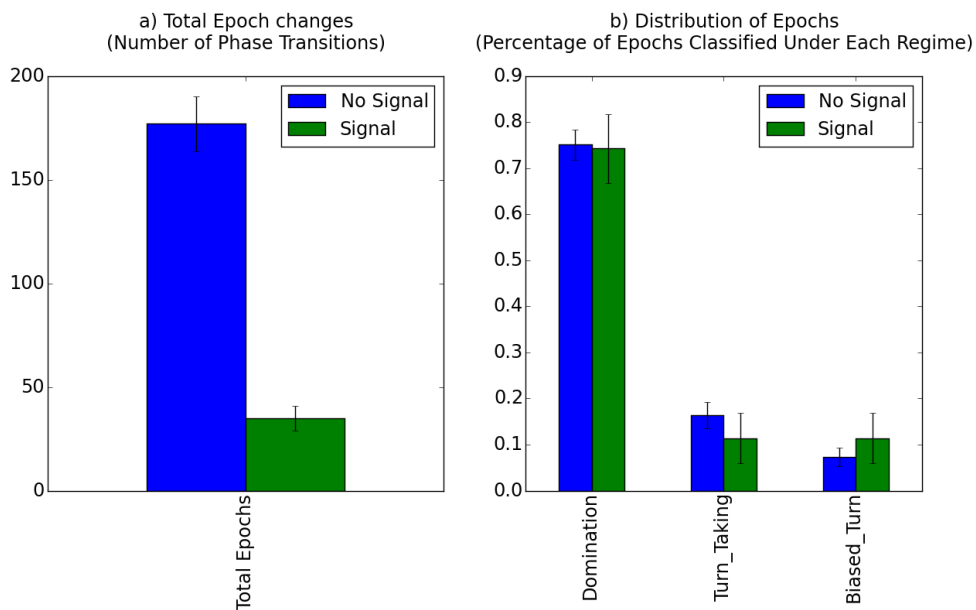


Figure 8: Panel a): Number of different epochs observed per treatment (i.e. time the system underwent a phase transition). Panel b): Distribution of epochs across regimes. Under the *Signal* treatment the system has less phase transitions (left panel), but the relative proportions across the regimes is the same for both treatments (right panel).

Figure 9 shows the average length of epochs per regime. As expected, the *Signal* treatment has longer epochs than *No-Signal*. But more importantly it also shows that Biased Turn-Taking has the longest epochs of all regimes for both treatments. So even if the number of Biased Turn-Taking epochs is low, their length makes it more frequent.

<sup>19</sup> Further tests on understanding better the difference in the frequency of transitions across treatments have been conducted (not reported), although preliminary results show that the causes might be quite complex. See section 6.3, on “future research”, for additional comments on this regard.

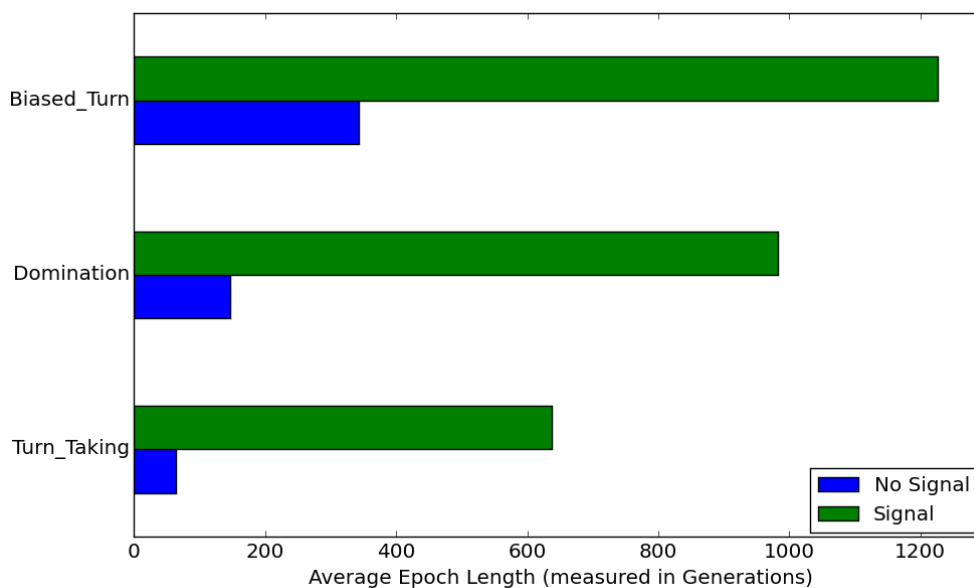


Figure 9: Average length of epochs per regime. Signal treatment presents the longest epochs.

Finally, it is worth emphasizing the difference in time spent under the Turn-Taking regime across treatments. In this case, the probability of a generation being classified as Turn-Taking goes from 11% without the signal to 22% when it is included (which can be seen in Figure 7)<sup>20</sup>. So, even if Turn-Taking is the least frequent regime, it is more likely to be found with the signal than without it.

Let us summarize these findings as follows:

**Result 3:** *Two main treatment effects are identified: first, the system undergoes fewer epoch changes under the Signal treatment: a total of 35 compared to 177 for No-Signal. Second, the probability of finding the system under a Turn-Taking regime increases with the Signal from 11% to 22%. However, Turn-Taking is the least probable regime for the system. The system spends most of its time under Biased Turn-Taking regimes, with such epochs being longer, rather than more frequent.*

This result rules out CE being more frequent than other equilibrium concepts such as pure Nash, but it doesn't say anything about agents actually following the signal. For this, the strategies that are being used under Turn-Taking epochs need to be evaluated in a different way, both at the macro and micro level.

#### 5.4 Searching for CE behavior at the aggregate level

In order to analyze behavior consistent with CE, the reader is reminded that, here, CE is being used only to refer to an equilibrium in which agents

---

<sup>20</sup> Differences here are statistically significant: since the sample is so large and the units of observation are generations, each one taken as an independent observation (with  $t=100,000$ ), standard errors of the mean are on the order of  $1 \times 10^{-5}$ , resulting in very small confidence intervals.



condition their actions by using the signal. This avoids using CE to describe other types of behavior such as pure Nash.

The first way to explore if there are Turn-Taking epochs in which agents are following the signal, is to develop an aggregate measure based on the probabilities of agents playing each action conditioning on the signal. The intuition is that if agents are following the signal, one should observe, on average, that the probability of playing the same action (say A) should be high when the same signal is observed (say Heads). On the contrary, if the signal is being ignored, one should not expect the same action for each signal. Although this measure doesn't show exactly how agents are coordinating (such analysis is done in section 5.5), it will allow us to identify if there are epochs in which the signal is consistently being followed.

Let  $p(\text{row} = A | S = \text{Tails})$  be the observed probability in a particular generation for an agent from population ROW to play A, given that the observed signal for that round was Tails. Then, using analogous notation for a player from population COL, action B and signal Heads, we define our *Correlated Equilibrium Measure* in generation  $t$  ( $CEM_t$ ) as follows:

$$CEM_t = \frac{p(\text{row} = A | S = \text{Tails})p(\text{col} = A | S = \text{Tails}) \times p(\text{row} = B | S = \text{Heads})p(\text{col} = B | S = \text{Heads}) + p(\text{row} = A | S = \text{Heads})p(\text{col} = A | S = \text{Heads}) \times p(\text{row} = B | S = \text{Tails})p(\text{col} = B | S = \text{Tails})}{2}$$

Notice that  $CEM_t \in [0,1]$ . If agents are using the signal,  $CEM_t \approx 1$ . Under a Domination regime,  $CEM_t \approx 0$ . If agents are Turn Taking but ignoring the signal,  $CEM_t$  will be somewhere in-between.

We find that the behaviour of the values of  $CEM$  are very stable within single epochs. Agents use the signal in the same way within epochs, meaning that within a single one, agents tend to use the signal in the same way. Appendix 7.3 shows the  $CEM$  values vs. the average payoffs for the *Signal* treatment.

Average <i>CEM</i>	Regime	Number of Epochs
0.12	Turn-Taking	2
0.86	Turn-Taking	2
0.11	Biased Turn Taking	1
0.23	Biased Turn Taking	1
0.40	Biased Turn Taking	2
0.00	Domination	26

Table 2: Average Correlated Equilibrium Measure (*CEM*) for all observed epochs under the *Signal* treatment. Calculated as the average  $CEM_t$  of all generations within a single epoch. Different epochs under the same regime can have the same average *CEM*, which is reflected in the “Number of Epochs” column.

Table 2 presents the average values of  $CEM_t$  for all different epochs observed under the *Signal* treatment. The values in the left column are the average  $CEM_t$  across all generations within a single epoch. Different epochs can have the same *CEM* value, which is shown in the “Number of Epochs” column. The first two rows of the table indicate that out of a total of four observed Turn-Taking epochs, the average value of  $CEM_T$  is 0.12 for two of them and 0.86 for the other two. Thus, in two of the Turn-Taking epochs agents are following the signal. This is our first evidence showing that agents have indeed learned to play CE. Yet, despite agents being able to learn coordination by using the exogenous signal, they can also ignore it completely and alternate as they would do in the absence of a signal<sup>21</sup>.

Unexpectedly, the *CEM* also shows that the behaviour under Biased Turn-Taking regimes can vary widely in its use of the signal. This behaviour will be explored below when analysing at the micro level the strategies that emerged, but it is worth mentioning that agents use the signal in different ways: this is what leads to the various observed intermediate values of the *CEM* in Table 2.

How important are the epochs where agents are learning to use the signal? The total time the system spends under a Turn-Taking regime in the *Signal* treatment is 22% (Figure 7), corresponding to four different epochs. However, the two epochs with a high average *CEM* constitute only 6.2% of the total time. So even if the evidence shows that agents can indeed learn to alternate their actions by following the signal, this happens rarely in the system<sup>22</sup>.

<sup>21</sup> Duffy et al. (2014) found similar results in their experiments. They document evidence in a BOS game where subjects exhibit both types of behaviour, alternating both by using the signal as well as by ignoring it.

<sup>22</sup> Why are agents not learning CE more often? One potential answer is that the learning algorithm is having difficulties in finding complex solutions (strategies) that include processing the signal. If the latter is true, one could argue that the results are driven by an inefficient algorithm instead of some deeper property of the system’s dynamics. Appendix 7.4 implements a test that addresses this issue. Results show that without the strategic component of the game, agents can easily learn to alternate their actions by using the signal.

**Result 4:** *Agents can learn to play CE and alternate their actions tied to an external signal. However, the likelihood of finding such behaviour is small. Agents can also learn to alternate by completely ignoring the signal. No evidence is found of CE being learned faster, or more frequently, than other types of behaviour.*

Thus while agents can indeed learn to play by conditioning on the signal, such learning occurs very rarely, and CE may not be the best descriptive notion of actual behaviour.

The one remaining question is related to how exactly are agents coordinating. Regimes and epochs classification hint at what agents are playing and gives us a characterization of the system at the macro level, but several different strategies at the micro level can lead to the same aggregate patterns. For example, even without the signal, Turn-Taking behaviour could be happening by playing (A,A) four times in a row followed by (B,B) four times, or by alternating one time on each. Understanding precisely what strategies have evolved is also important for the Biased Turn-Taking regimes. Not only does the system spend most of its time under such epochs, but the different values observed for CEM suggest that coordination happens under a wide range of behaviours. Such heterogeneity is impossible to grasp based on the aggregate measures presented so far as exploring such findings requires a more fine-grained micro analysis of what strategies evolved under each regime.

## 5.5 Micro Analysis

### 5.5.1 Individual Machines

Here we observe the exact structure of the most successful strategies playing under each regime. How are strategies responding to both the signal and the rival's actions? One first approach to understand these micro characteristics of the agents is to observe the top evolved individual machines.

Figure 10 shows some of the most frequent machines for each regime, chosen by randomly picking one epoch and selecting the most frequent strategy in one population<sup>23</sup>. The most frequent machine for one Domination epoch (Column population) is shown in panel (a), showcasing a very simple kind of behavior: play A no matter what. Perhaps surprisingly, simple strategies can perform very well in complex environments (see for example Gigerenzer et al. (2002) or Gigerenzer et al. (2011)). Strategies for Turn-Taking and Biased Turn-Taking are a bit more complex, but still far away from using all eight states. Even so, it becomes difficult to gain a clear insight about the system by observing only individual strategies. For example, it is hard to infer directly from the Turn-Taking machine (panel (b)) if that strategy follows the signal. For some particular cases (such as the automaton in panel (b) of Figure 3) this can be easier, but in general it is not trivial.

---

<sup>23</sup> The reader is reminded that the machines all have eight internal states, but that some of those states can be inaccessible or redundant (e.g. a machine with all eight states having an action of A has the same behaviour as a machine with one single state with action A). The shown machines are the minimal equivalents.

To be able to make such inferences one often needs to observe also the opponents' strategies. These are shown in Figure 11 for the Domination and Turn Taking regimes.

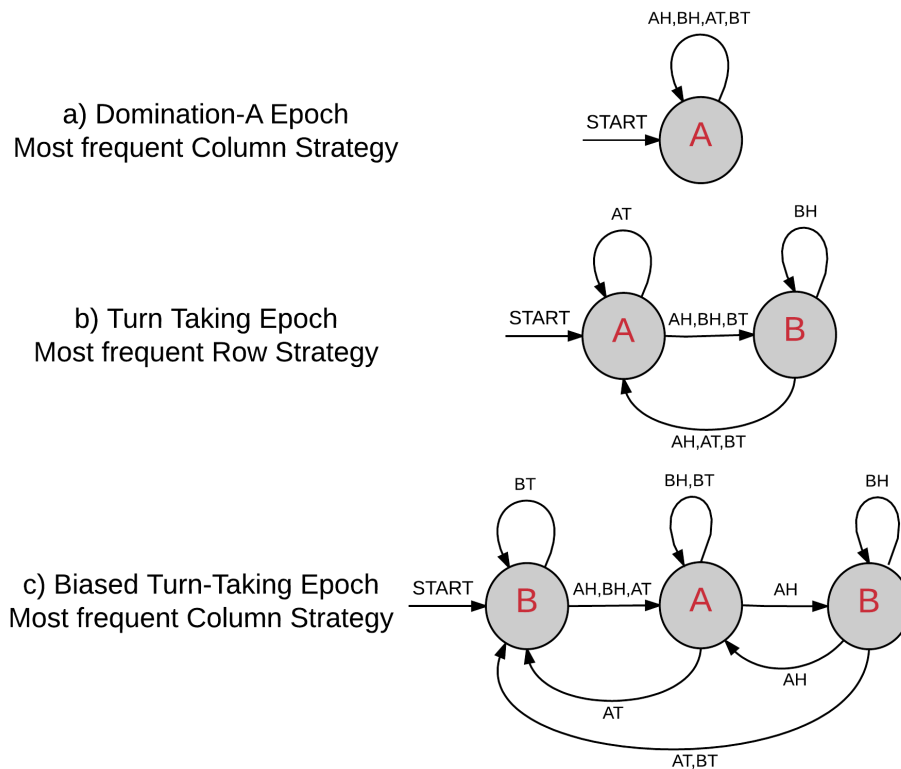


Figure 10: Some evolved strategies from Signal treatment. These were chosen by randomly selecting an epoch of the corresponding regime and choosing the most frequent strategy for one of the populations.

As can be seen, the top (most frequent) strategies in the opposing population are more complex. By observing the two interacting machines in panel (b) (of both Figure 10 and Figure 11), it is difficult to infer if they are following the signal or not. How exactly are they managing to coordinate?<sup>24</sup> So directly observing the strategies may not be the best way to analyze the system at the micro level, unless one limits the strategies to a few internal states.

Another way to analyze the machines, previously used in the literature (e.g. Miller (1996) or Ioannou (2013)) is to generate average measures based on the accessible states of the machines. For example, checking how many of the accessible states in each machine have particular behavioral traits has been used to describe cooperation games (e.g. how many states punish defections, or how many forgive one).

Although this approach has proven very useful, it doesn't come without limitations. To illustrate this, observe that larger strategies may not necessarily

<sup>24</sup> These two strategies, when playing against each other, actually do follow the signal.

use all of their states even if they are accessible<sup>25</sup>. A machine could only visit a subset of the accessible states if no rival machine gives it the necessary input. So focusing the analysis on measures of the states of the individual machines can be misleading, because it could include behavior that is never actually used.

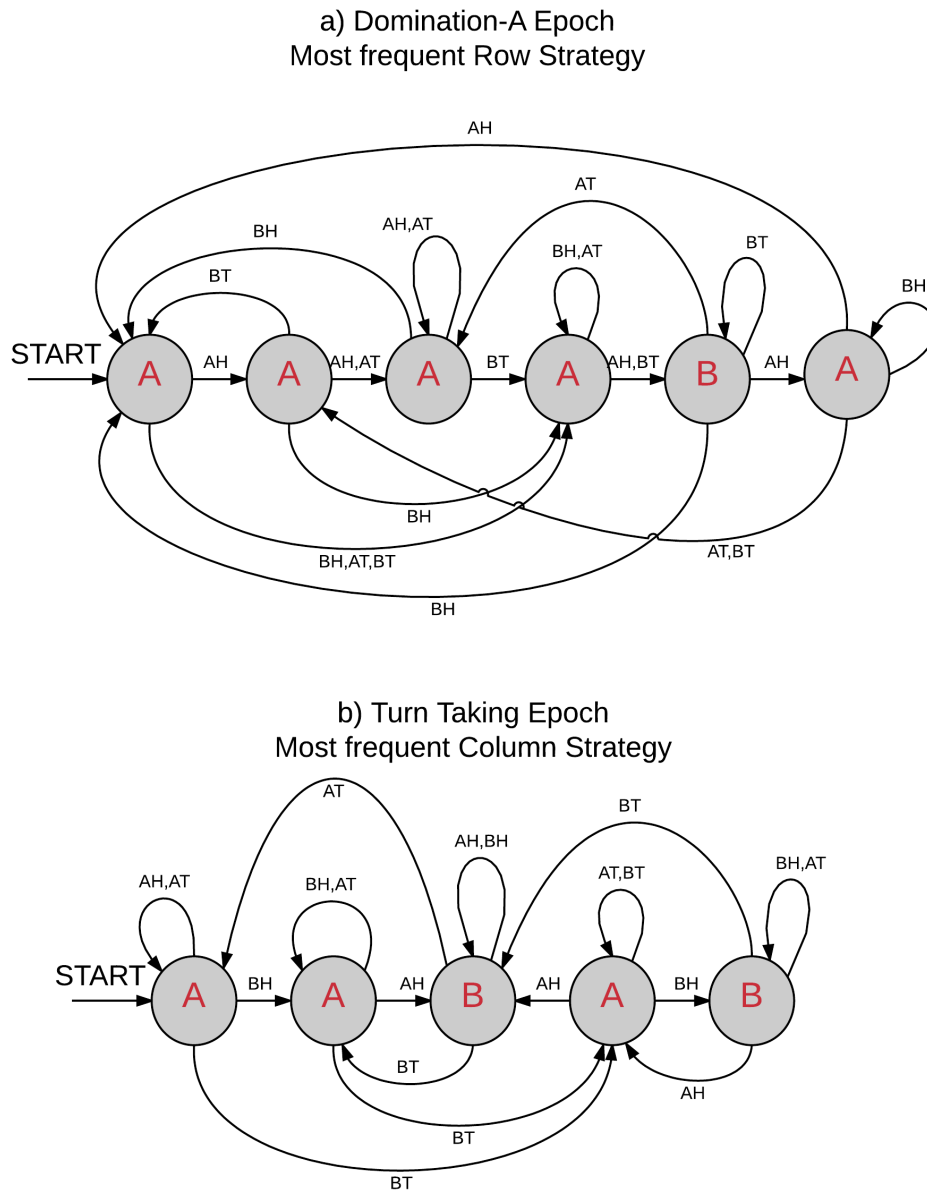


Figure 11: Evolved complex individual strategies. They were chosen by randomly selecting an epoch of the corresponding regime and choosing the most frequent strategy for one of the populations. The strategies presented are considered among

<sup>25</sup> As a reminder, a state is accessible if there exists at least one combination of inputs (i.e. opponent's last action and exogenous signal) that can lead the machine to be in that state.

*the complex ones (i.e. having more internal states). Is very difficult to infer the behavior of the system by observing them.*

In summary, focusing the micro behavior on the analysis of individual machines presents two potential difficulties: first, single machines don't capture the interaction between strategies. Second, average measures of the accessible states can be misleading, for not all of them are necessarily visited. So how can such analysis be done? In order to solve these issues, this paper uses 'Joint Machines' analysis<sup>26</sup>.

### 5.5.2 Joint Machines

The interaction between any two automata can be modeled as a Joint Machine (JM). A JM is a 'meta' machine that represents, in a single automaton, the observed behavior of two automata playing each other. An example is appropriate to understand it.

Figure 12, in panels (a) and (b) shows automata for the *No-Signal* treatment. It is not straightforward to understand how they are coordinating by directly observing them, but panel (c) shows the corresponding JM. Both interacting machines start playing B in their initial state, which is represented by a starting state of the JM with action BB. The machine in (a), after observing B, transitions to its last state with action A, while the machine in (b) transitions to a state with action B (also its last). These actions are captured by the second state of the JM, with a joint action of AB. Following the same logic, using the input received by each machine and the state they transition into, the JM captures the actions in states that are visited. In Figure 12, by observing the JM in (c), it is easy to notice that after the two initial rounds, both machines will take turns, alternating their coordination point from AA to BB and back to AA, indefinitely. These machines correspond to a Turn-Taking regime.

Notice that JMs' actions are no longer the action of one particular strategy, but those of both interacting machines that are being represented. A state of the JM is given by corresponding states of the two interacting machines. So if the action of the JM is, for example AB, it means that in that particular state one agent plays A and the other B. This representation makes a JM a simpler representation of complex behavior.

---

<sup>26</sup> This approach is an original idea of, and has been developed by, professor John H. Miller. The implementations here are based on his own original algorithms via personal communication.

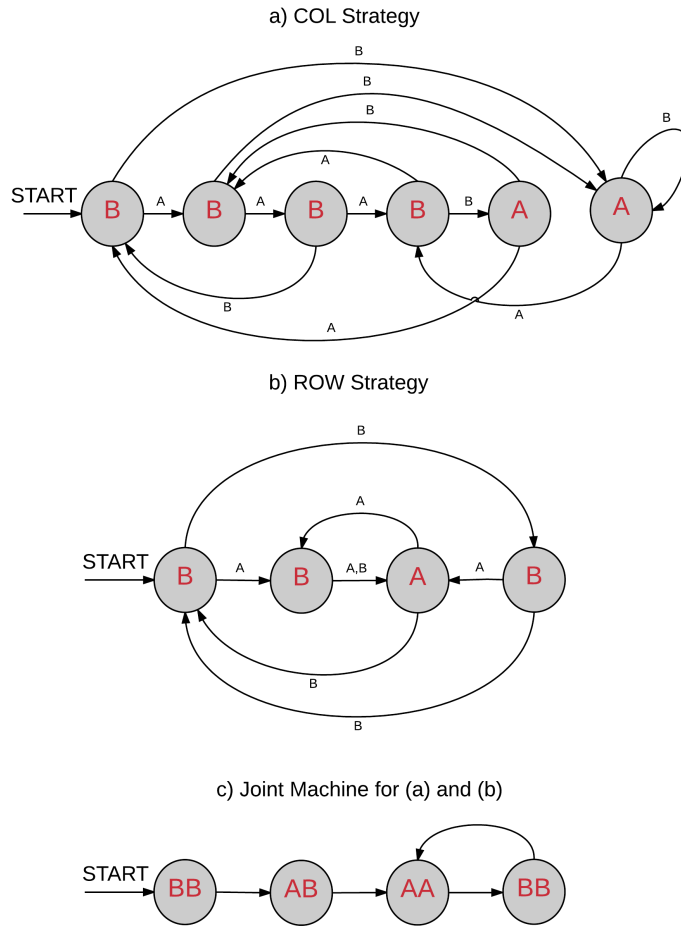


Figure 12: Example of Joint Machine for No-Signal treatment. When the machines in (a) and (b) play each other, their interaction can be represented as the Joint Machine in (c). These machines evolved under a Turn-Taking regime.

Without the signal there is no stochastic component, so the JM is completely deterministic (as the one in panel (c) of Figure 12). With the signal, transitions of the JM will depend only on the stochastic observed signal  $H$  or  $T$ . In any case, since the constituent automata are finite, the JM at some point will return to one state-pair that has already been visited, and from there cycle between a subset of states indefinitely<sup>27</sup>. The focus in what follows of this section is on the *Signal* treatment, but some intuition about JMs under *No-Signal* can be found in Appendix 7.5.1.

<sup>27</sup> In the formal definition of automata in section 4.2, the following are the differences when the automata defined is a JM instead of a single strategy. For both *Signal* and *No-Signal* treatments, the JM actions are  $A_i \in \{AA, BB, AB, BA\}$ . For *Signal*, now  $W = S \in \{H, T\}$ , meaning that the machine no longer depends on the input  $A_{-i}$  (opponent's action last round) since such information is already contained in the actions of each internal state. Under *No-Signal* the machines are simpler:  $W$  is no longer defined since the JM doesn't depend in any input or state of the world. The transitions are deterministic with  $\tau: Q \rightarrow Q$ , with each state  $Q$  having one single transition into another  $Q$ .

### 5.5.2.1 Turn-Taking Joint Machines

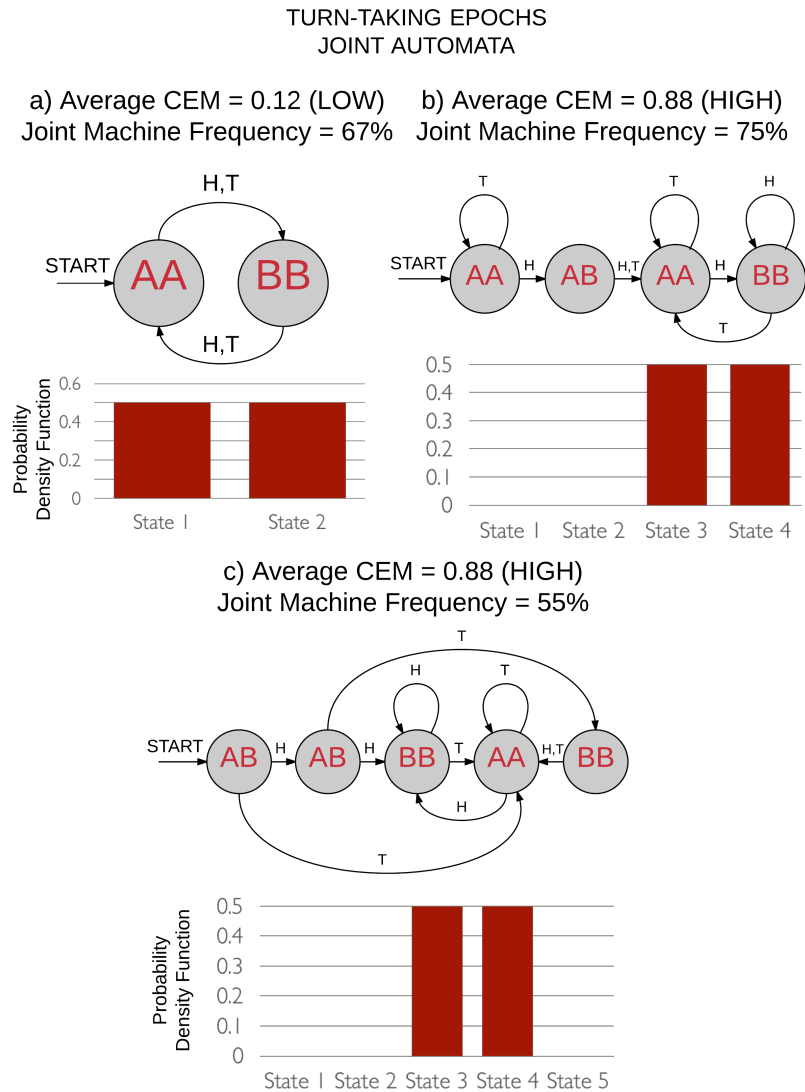


Figure 13: Joint Automata that evolved during the Turn-Taking epochs of the Signal treatment. Each machine was picked from the corresponding epoch (with low or high CEM value). One generation was randomly chosen from that epoch, and the most frequent Joint Machine is the one shown. Probability Density Functions show the long-run probability of finding the machine in each particular internal state.

Observe the JM presented in panel (a) in Figure 13, which is one of the three representative JMs shown for three different Turn-Taking epochs. The machine has only two states. In the starting one, both machines play A, and whatever the observed signal is (either H or T), it will always transition into the second state. In the second state, the action is BB, and again the transitions are the same regardless of the signal, returning into the initial state. This JM represents two strategies that when interacting will take turns playing (A,A), then (B,B), then (A,A) and so on. Notice that this machine completely ignores the signal, but still manages to perfectly alternate. This is precisely what the CEM captures at the aggregate level. Such machines belong to a Turn-Taking



regime with  $CEM_T = 0.12$ : a low value reflecting that under such regime agents are not relying on the signal to coordinate their actions.

The JMs presented in Figure 13 are representative of the behavior observed during each epoch<sup>28</sup>. They were chosen by randomly picking a generation from an epoch with the corresponding CEM and then selecting the most frequent JM. For example, for the machine in panel (a), its frequency is 67%. This means that 67% of all the pairs of strategies playing each other in such generation are described by this automaton<sup>29</sup>.

Associated with each machine, there is a Probability Density Function (PDF). It shows the probability of finding the machine in each state in the very long run. States that have zero probability would only be visited before the JM starts cycling, so in the long run their probability tends to zero. Those states with positive probabilities are the ones characterizing the core behavior of the system, and will be referred to as the *cycling states*. Finally, it is worth noting that the cycling states are also very stable across epochs. Even if the JMs don't represent 100% of the interactions, usually the states in the cycle do. Two JMs can have different states before reaching the cycle, but once there, their behavior is very similar. This is the case for JMs in panels (b) and (c), having different states with low probability, but the same cycle. JMs, and particularly the states with positive probabilities in the PDF, are an excellent tool for understanding the micro behavior of the system.

Let us also explain the behavior found in panels (b) and (c). Such JMs give us another formal way to understand the CE learned by the agents. Notice the cycling states (again, the ones with positive probability in the PDF). Even if both machines are from different epochs and have different states, their cycling behavior is identical. In both JMs the behavior alternates between AA and BB depending on the signal: in any of the two cycling states, whenever the signal is T, it will transition to the actions AA. Whenever it is H, it will transition to actions BB. This shows that the machines have learned to interpret the signal and coordinate based on it. As expected, on average, most JMs found under Turn-Taking epochs (the three panels) will play 50% of the times AA and 50% of the times BB. The difference —what is being captured by the CEM—is whether their transitions depend on the signal or not. This can be easily grasped in the JMs by observing the transitions in the cycling states.

---

<sup>28</sup> A total of four Turn-Taking epochs were identified for the *No-Signal* treatment. Only three machines are shown because the JM that doesn't follow the signal (panel (a)) was found to be representative under two of them. The other two epochs with high CEM values are shown in order to highlight that even if the machines are different, their core behaviour can be the same.

<sup>29</sup> With 40 agents in each population, the total number of possible Joint Machines in each generation is  $40 \times 40 = 1,600$ .

### 5.5.2.2 Biased Turn-Taking Joint Machines

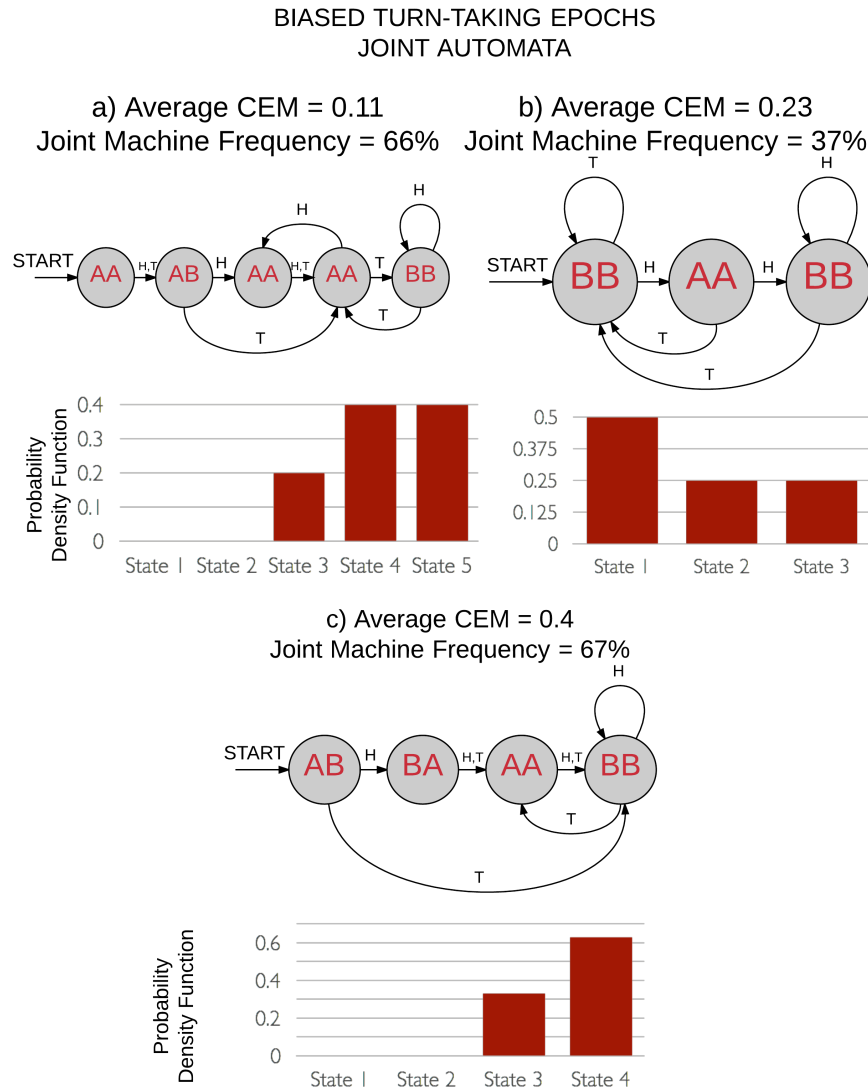


Figure 14: Joint Automata that evolved during a Biased Turn-Taking epoch of the Signal treatment. Each machine was picked from an epoch having a different CEM value. One generation was randomly chosen from that epoch, and the most frequent machine is the one shown. Probability Density Functions show the long-run probability of finding the machine in each particular internal state.

The corresponding analysis for the Biased Turn-Taking regime is also presented. Representative JMs for epochs with different CEM values are shown in Figure 14. Notice that the ratio of AA to BB actions varies across JMs (observe the probabilities of each machine being in AA or BB during the cycling states). This characteristic is impossible to grasp by observing only the aggregate classification based on the regime.

Each machine uses the signal differently in each state. For example, the JM in panel (a) interprets the signal consistently in state 4 and state 5, but not in

its other states. In state 4, it transitions to AA given H and to BB given T. This means that the two constituent strategies found a way to coordinate by using the signal in that particular state. In state 5 an interpretation to the signal is also given, but it is the opposite of that in state 4: BB when H and AA when T. In state 3, the machine completely ignores the signal and always transitions to state 4. Here again, such behavior would have been impossible to observe based only on the aggregate CEM measure of 0.11 (and very difficult to grasp based on the individual machines). The JM’s cycling states and the associated PDFs allows an understanding of how partial signal following is happening. Similar intuitions can be made for the other JMs in the Biased Turn-Taking epochs.

Perhaps the reader could have had an accurate a priori intuition of the kind of behavior observed under the Turn-Taking regime based on the values of CEM. Even so, the JMs present a much more intuitive and clear analysis of how machines coordinate. But for the Biased Turn-Taking epochs, such a priori expectations are more unlikely: the varied and perhaps less intuitive ways in which agents follow the signal were not hypothesized and were surprising. Potentially finding *some* strategies that follow the signal and others that don’t was initially thought of. But observing such behavior under one single interaction (one single pair of agents) that represents strategies able to follow the signal or ignore it at the *same* time, was unexpected. This is a nice example of how adaptation can come up with marvelous and unexpected solutions that would be difficult to anticipate.

*Result 5: Analysis based on Joint Machines (which summarizes any two interacting strategies) is more clear and robust than analyzing individual machines. For the Turn-Taking epochs, such analysis shows how some agents completely ignore the signal and how others perfectly condition on it. For the Biased Turn-Taking epochs, it shows that single machines can at the same time ignore, partially use, or perfectly follow the signal depending on the history of the game (i.e. their internal state).*

## 6 CLOSING REMARKS

### 6.1 Summary

This paper uses an explicit evolutionary process, simulated by a genetic algorithm, in order to analyze the effects of an exogenous signal in a repeated Battle of the Sexes coordination game. Its focus is on analyzing the strategies that emerge when coordinating with boundedly rational agents.

Contrary to what was expected, with and without the signal, coordination behavior was quite similar, presenting the same types of equilibria (such as pure Nash and alternation, both symmetrical and asymmetrical). Interestingly, the system doesn’t settle down to a single equilibrium, but rather exhibits a constant transition from one to the other. Efficiency in terms of payoffs is the same with and without the signal, meaning that agents coordinate equally well under both setups. The main difference found was the frequency of transitions from equilibrium to equilibrium: with the signal, there is more stability (i.e. less transitions).

Our adaptive agents can indeed learn to coordinate consistently using the signal as a “recommendation” of what to play. However, such behavior is

learned very rarely (around 5% of the time), making other strategies a more likely descriptive notion of observed behavior.

This is the first work using adaptive agents in the long run to study coordination games that include a signal. The above results constitute our main findings regarding Correlated Equilibrium. It also analyzes automata by focusing on Joint Machines: ‘meta’ automata that summarizes several interacting agents in a single representation. This analysis permitted us to analyze complex agents. Regarding Correlated Equilibrium, this analysis showed that when the signal is included *some* strategies can alternate their actions by using the signal while *others* can do so by completely ignoring the signal. It also allowed us to see that signal interpretation is not necessarily as intuitive as one might think, and that complex strategies learned to use signals in very different ways. For example, the *same* strategy can, depending on the history of the game, sometimes use the signal as different “recommendations” of play, partially use it, or completely ignore it. This complexity on how strategies use the signal would be difficult to observe without this methodology.

Previous studies of signal use in coordination game have mainly been done with experiments. Conclusions from such experiments show that even though some subjects can indeed learn to alternate their actions by following an exogenous signal, this rarely happens, as they can also alternate by completely ignoring signals (as in Duffy and Feltovich (2010)). One of the main advantages of evolutionary simulations is that they allow agents to learn over considerably longer time spans than in the lab. The limited time span of the lab has led some authors (e.g. Cason and Sharma (2007)) to speculate that signal conditioning would probably be learned much more often if agents were given more time. Our results, however, show that this is not necessarily the case, reinforcing previous results that cast doubt on the notion of Correlated Equilibrium as an accurate description of commonly observed behavior.

## 6.2 Discussion

The Battle of the Sexes game has both coordination and conflict elements (Camerer (2003), p.354; Lau and Mui (2008), p.154). This “mixed motive” social situation arises because both players want to coordinate and choose the same action (a social or shared motive) but also disagree on the activity they want to coordinate on (an individual motive). Our results show that the coordination dimension is solved most of the time, with or without the signal: the system is equally efficient most of the time. But the degree of conflict inherent in the solutions (equilibria) found by the agents can vary at different moments in time. When agents are taking turns symmetrically, they have found a solution without any conflict in terms of received payoffs, but when playing one of the Nash solutions consistently or under asymmetric turn taking, the conflict dimension is not solved. We can make a distinction in the behavior of the system in terms of the time span analyzed. In the short run, coordination seems to dominate over conflict. But since regimes are subject to change and transitions, in the (very) long run the conflict issues are averaged out. So in the long run the system has both coordination an absence of conflict (or efficiency and equality), but at any moment in time only coordination is found for sure.

Regarding the effects of the signal, in theory it could help agents solve both coordination and conflict. However, since agents learn to coordinate quite well

without it, the signal is addressing a problem that doesn't need help to be solved. The signal could also solve the conflict dimension, but in evolutionary terms, it only does that in occasion according to our model. At the heart of this distinction, is the game theoretic induction approach to solve these problems. Theoretically, agents could reason a priori and arrive to a common understanding about how to use the signal to solve both coordination and conflict, hence playing conditioning on the signal. But this would require a lot of reasoning and common knowledge. And notice that such outcome is only one possibility consistent with traditional rationality, since one agent being completely stubborn and only playing its preferred action, with the other complying, is a Nash Equilibrium.

In summary, it seems that the signal doesn't have the expected effect in behavior because agents don't really need it to coordinate. And even if the signal could solve the conflict dimension in the short run, the system can still operate under different degrees of conflict, since it doesn't lead to miscoordination or efficiency losses. In the long run, without the signal, both dimensions are solved, so the introduction of the signal seems redundant.

### 6.3 Future research

One of the main behavioral differences found between the *No-Signal* and *Signal* treatments was the difference in number of transitions. Tests on alternative treatments have been conducted, hinting that such results can be related to how the mutation rates interact with the number of transitions in the machines. However, results are not conclusive. The problem seems more complex than anticipated, requiring the development of better performing software than the one currently being used. Not only being able to run simulations for longer time spans could aid in this regard (which would reduce potential effects of very long epochs), but would also allow more efficient exploration of other potential variables that could also be related<sup>30</sup>.

Answering the above is also related to more general questions, to be pursued in the mid and long-term. Recent efforts in evolutionary biology have focused on understanding similar phase transitions in natural systems, and other areas ranging from statistical physics, to artificial life to evolutionary robotics, have already made some contributions in understanding general principles of such changes across domains<sup>31</sup>. The computational nature of our model makes detailed analysis of all its components feasible, at least in principle. Understanding what mutations at the micro level are necessary for the system to transition, what aggregate measures show that the system is "ripe" for a sudden change and what precise evolutionary pathways are followed when this happens, will certainly shed some light not only in better understanding equilibrium behavior in systems with boundedly rational

---

<sup>30</sup> Several of this tools have already been implemented. Some measures such as evolutionary "waste" or inefficiency in the construction of the machines, or unused behaviour related to unvisited states present in the machines (reflecting potential for change in the system) have already been explored. However, their examination is currently very expensive in computational terms, requiring further development on the implemented software.

<sup>31</sup> Solé (2016) presents a recent review of contributions across different fields. Sornette (2004) is an example of how understanding phase transitions is relevant for social sciences, in this case financial markets.

agents, but also into understanding phase transitions in evolutionary, artificial and social systems.

## 7 APPENDIX

### 7.1 Formal presentation of correlated equilibrium and the one shot BOS game<sup>32</sup>

#### 7.1.1 Correlated strategy pairs: relations with pure and mixed strategies

In a game one shot game with two players having two possible actions the general form of a correlated strategy pair is

		Player 2	
		C	D
Player 1	A	$p_1$	$p_2$
	B	$p_3$	$p_4$

where  $p_1 + p_2 + p_3 + p_4 = 1$ . Such strategy can be represented as a 4-dimensional vector  $\pi = (p_1, p_2, p_3, p_4)$ , meaning that (A,C) is played with probability  $p_1$ , (A,D) is played with probability  $p_2$ , etc.. Under correlated strategy  $\pi$  the expected payoffs or rewards of player  $i$  are denoted as  $R_i(\pi)$  and calculated with respect to the joint distribution of the actions to be taken. So such payoffs are given by a linear combination of the  $p_i$ :

$$R_i(\pi) = p_1 R_i(A, C) + p_2 R_i(A, D) + p_3 R_i(B, C) + p_4 R_i(B, D)$$

Notice the relationship between a correlated strategy pair and other strategy types. If  $p_i = 1$  for some  $i$ , then the correlated strategy pair is a pair of pure strategies. If  $\pi$  is of the form  $(qr, q[1 - r], [1 - q]r, [1 - q][1 - r])$  then it corresponds to a pair of mixed strategies. Here, Player 1 takes action A with probability  $q$  and Player 2 takes action C with probability  $r$ , with such probabilities being independent of the action of the rival. This makes the set of correlated strategy pairs an extension of the set of mixed strategy pairs.

In general, to attain a correlated strategy pair communication is required, with an agreement on it before the game is played. However, the agreement is not (and cannot be made) binding, so players are free to ignore any recommendation.

#### 7.1.2 Conditions for a CE in a 2x2 matrix game

According to strategy pair  $\pi = (p_1, p_2, p_3, p_4)$ , Player 1 is recommended (by the randomization device or the external third party) to play A with probability  $p_1 + p_2$ . Given that Player 1 is recommended to play A, the probability of Player 2 being recommended to play C is  $\frac{p_1}{p_1 + p_2}$ .

---

<sup>32</sup> A textbook presentation on correlated equilibrium can be found in Myerson (1997). The one in this appendix was greatly benefited from the lecture notes of Dr. David Ramsey used at the University of Limerick, found online at [http://www3.ul.ie/ramsey/Lectures/Operations\\_Research\\_2/gametheory4.pdf](http://www3.ul.ie/ramsey/Lectures/Operations_Research_2/gametheory4.pdf) (last visited on April 25 of 2016).

In a CE each player should maximise her expected payoffs  $R_i(\pi)$  given the recommendation (signal) she receives. So if Player 1 is recommended to play A, her expected payoffs under such a correlated strategy pair are

$$R_1(\pi) = \frac{p_1 R_1(A, C)}{p_1 + p_2} + \frac{p_2 R_1(A, D)}{p_1 + p_2}$$

If Player 1 ignores her recommendation to play A and she plays B instead, her expected payoffs are

$$R_1(\pi) = \frac{p_1 R_1(B, C)}{p_1 + p_2} + \frac{p_2 R_1(B, D)}{p_1 + p_2}$$

For stability it is required that

$$\frac{p_1 R_1(A, C)}{p_1 + p_2} + \frac{p_2 R_1(A, D)}{p_1 + p_2} \geq \frac{p_1 R_1(B, C)}{p_1 + p_2} + \frac{p_2 R_1(B, D)}{p_1 + p_2}$$

which leads to

$$p_1 R_1(A, C) + p_2 R_1(A, D) \geq p_1 R_1(B, C) + p_2 R_1(B, D)$$

For the sake of completion, notice that the above expression is not defined in the case where  $p_1 = p_2 = 0$  since we would be dividing by zero. However, in this case Player 1 is never recommended to play A and this condition might then be ignored.

The same line of argument given above can be used for the conditions corresponding to the following recommendations: i) Player 1 to play A, ii) Player 1 to play B, iii) Player 2 to play C and 4) Player 2 to play D.

Hence, the four condition for a correlated equilibrium, respectively for the above recommendations are:

$$p_1 R_1(A, C) + p_2 R_1(A, D) \geq p_1 R_1(B, C) + p_2 R_1(B, D)$$

$$p_3 R_1(B, C) + p_4 R_1(B, D) \geq p_3 R_1(A, C) + p_4 R_1(A, D)$$

$$p_1 R_2(A, C) + p_3 R_2(B, C) \geq p_1 R_2(A, D) + p_3 R_2(B, D)$$

$$p_2 R_2(A, D) + p_4 R_2(B, D) \geq p_2 R_2(A, C) + p_4 R_2(B, C)$$

There are some relationships between correlated equilibria and other types of equilibria that are worth mentioning. First, any Nash equilibrium pair of strategies is also a correlated equilibrium. Second, a pair of mixed strategies that is not a Nash equilibrium is not a correlated equilibrium. Third, any randomization over Nash equilibria is also a correlated equilibrium. Finally, any randomization over a set of strong Nash equilibria can be attained by joint observation of a public signal<sup>33</sup>.

### 7.1.3 Battle of the sexes correlated equilibrium

The CE solution of interest in this paper for the BOS game, as indicated in the main text, is the one given by a fair coin toss as the exogenous signal. Formally, such CE is described as  $\pi = \left(\frac{1}{2}, 0, 0, \frac{1}{2}\right)$ . This equilibrium is both an

---

<sup>33</sup> The two last conditions allow the easy graphical representation of the convex hull for correlated equilibria, as in the main text. In such case for the BOS, both (A,C) and (B,D) are strong Nash equilibrium, so any correlated strategy pair that picks (A,C) with probability  $p$  and picks (B,D) otherwise, is a correlated equilibrium.



*utilitarian* and an *egalitarian* equilibrium. Let us define such properties formally and then use the concrete payoffs examined in this paper in order to derive such solution.

- 1) **Utilitarian equilibrium:** an equilibrium which maximizes the sum of the expected payoffs of the players
- 2) **Egalitarian equilibrium:** an equilibrium which maximizes the minimum expected payoff of a player.

Since the expected payoff of players are linear combinations of  $p_i$ , the criteria above can be expressed as a maximization of a linear combination of  $p_i$ . So equilibria of such types can be derived by defining the problem as a linear programming one. For this, consider the following payoff matrix, with the same rewards of interest as in the main text:

		Player 2	
		A	B
Player 1	A	2,3	0,0
	B	0,0	3,2

The utilitarian equilibrium can be found by solving the following problem:

$$\max z = (2 + 3)p_1 + (0 + 0)p_2 + (0 + 0)p_3 + (3 + 2)p_4 = 5p_1 + 5p_4$$

subject to

$$p_i \geq 0 \text{ for } i = 1,2,3,4$$

$$p_1 + p_2 + p_3 + p_4 = 1$$

$$2p_1 + 0p_2 \geq 0p_1 + 3p_2 \Rightarrow p_1 \geq \frac{3p_2}{2}$$

$$0p_3 + 3p_4 \geq 2p_3 + 0p_4 \Rightarrow p_4 \geq \frac{2p_3}{5}$$

$$3p_1 + 0p_3 \geq 0p_1 + 2p_3 \Rightarrow p_1 \geq \frac{2p_3}{5}$$

$$0p_2 + 2p_4 \geq 3p_2 + 0p_4 \Rightarrow p_4 \geq \frac{3p_2}{2}$$

The first two restrictions represent the conditions for  $(p_1, p_2, p_3, p_4)$  to define a joint distribution. The final four conditions are the ones required for the solution to be a correlated equilibrium (as defined before).

One could solve this problem for all  $p_i$ , but there is a simpler way if one knows the pure Nash equilibria for this problem. Here, (A,A) and (B,B) are Nash equilibria that maximize the sum of the payoffs to the players over the set of pure strategy pairs. And any randomization over these two Nash equilibria is a correlated equilibrium that gives the same sum of payoffs. Hence, any  $\pi$  of the form  $\pi = (p, 0, 0, 1 - p)$  is a utilitarian equilibrium. So  $\pi = (\frac{1}{2}, 0, 0, \frac{1}{2})$  is a utilitarian equilibrium.

Let's turn now to the egalitarian equilibrium. For this, it is convenient to notice that the BOS game is not symmetric but still has a degree of symmetry. A 2x2 game where both players can choose either action A or action B will be called *quasi-symmetric* if the following conditions hold (which is indeed the case for BOS):

$$R_1(i, j) = R_2(j, i)$$

$$R_1(i, i) = R_2(j, j), \text{ where } i \neq j, \text{ and } i, j \in \{A, B\}$$

In words, this means that a payoff vector on the leading diagonal is the reverse of the other payoff vector on that diagonal.

As a result, at an egalitarian equilibrium of a quasi-symmetric game both players must obtain the same expected payoffs. So to find an egalitarian equilibrium of a quasi-symmetric game, the problem is to maximize the expected sum of the payoffs using the same constraints as before, but adding a new one: that both players should obtain the same payoffs. Hence the problem, is

$$\max z = 5p_1 + 5p_4$$

subject to (as before)

$$p_i \geq 0 \text{ for } i = 1, 2, 3, 4$$

$$p_1 + p_2 + p_3 + p_4 = 1$$

$$2p_1 + 0p_2 \geq 0p_1 + 3p_2 \Rightarrow p_1 \geq \frac{3p_2}{2}$$

$$0p_3 + 3p_4 \geq 2p_3 + 0p_4 \Rightarrow p_4 \geq \frac{2p_3}{5}$$

$$3p_1 + 0p_3 \geq 0p_1 + 2p_3 \Rightarrow p_1 \geq \frac{2p_3}{5}$$

$$0p_2 + 2p_4 \geq 3p_2 + 0p_4 \Rightarrow p_4 \geq \frac{3p_2}{2}$$

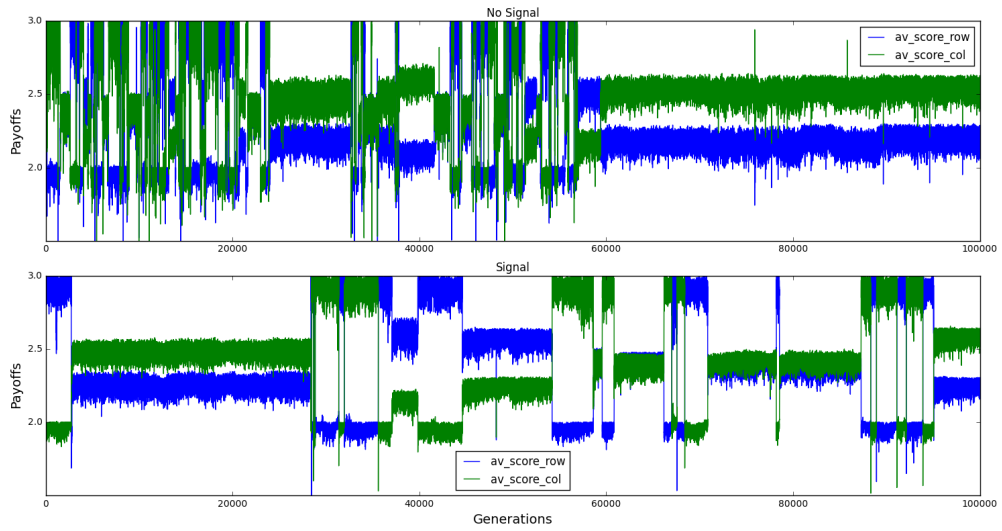
and adding the condition

$$2p_1 + 3p_4 = 3p_1 + 2p_4 \Rightarrow p_1 = p_4$$

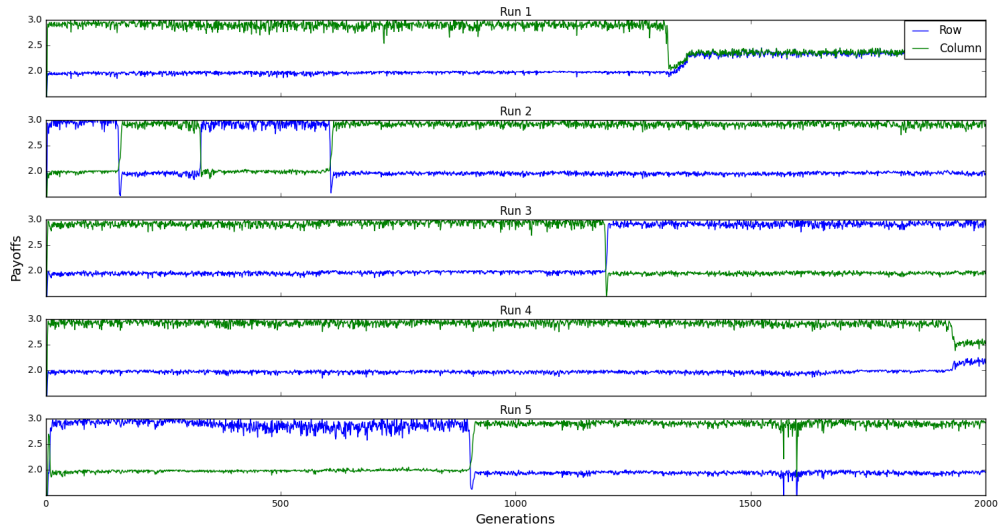
As before, any correlated equilibrium of the form  $(p, 0, 0, 1 - p)$  maximises the sum of expected payoffs. And observing that setting  $p = \frac{1}{2}$  holds for that new last condition, one can then define  $(\frac{1}{2}, 0, 0, \frac{1}{2})$  as the egalitarian equilibrium of interest.

## 7.2 Additional overview of average payoffs

Statistical analyses of the model are based on the simulations presented on Fig. A. One very long run with 100,000 generations is run for each treatment. Fig. B presents five shorter simulations for the *Signal* treatment, analogous to the figure presented on the main text for *No-Signal*.



*Fig. A: Average payoffs per population. Longest simulations for both Signal and No-Signal treatment with 100,000 generations. Statistical analyses in the main text are based on these runs of the model.*



*Fig. B: Signal treatment. Average payoffs per population. Each panel is one different run of the model, each consisting of 2,000 generations.*

### 7.3 Correlated Equilibrium Measure (CEM)

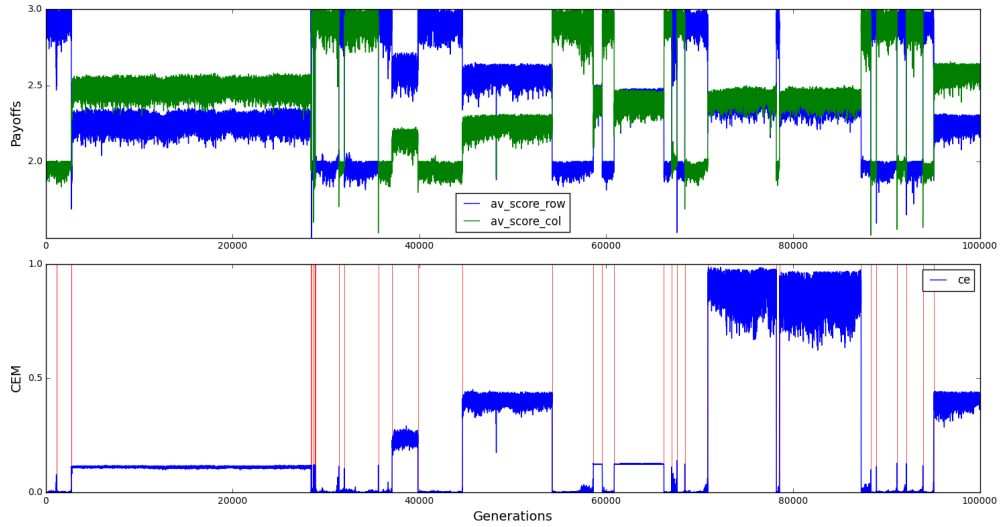


Fig. C: Average Payoffs on Signal treatment (top panel) vs. Correlated Equilibrium Measure CEM (bottom panel). Vertical lines in the bottom panel indicate the end of an epoch. It can be seen that within single epochs, the CEM is quite stable, meaning that agents use (or not) the signal in the same way consistently under each single epoch.

### 7.4 A learning test

Why agents don't learn to play CE more often? If the cause is that the algorithm finds it difficult to explore the larger strategy space when the signal is included, then not finding CE more often wouldn't be due to an interesting feature of the strategic interactions of the agents, but rather to having an inefficient (or perhaps wrongly 'tuned') learning mechanism. In order to address this concern, a test for the model in the *Signal* treatment was run by changing the payoffs of the game. The test is implemented by modifying the payoffs depending on the outcome of the signal in each round as follows:

<b>Payoffs if Signal = Heads</b>				<b>Payoffs if Signal = Tails</b>			
		Player 2				Player 2	
		A	B			A	B
Player 1	A	3,3	0,0	Player 1	A	0,0	0,0
	B	0,0	0,0		B	0,0	3,3

Notice that with these payoffs there's no conflict of interests between the agents. If they are able to follow the signal in this environment, it means the algorithm is not having difficulties exploring the larger strategy space (compared to *No-Signal*). The model was run five different times up to 2,000 generations. Under this setup both generations will have the exact same payoffs and the Pareto optimal is now three for both populations.

In all of the simulations agents quickly learned to follow the signal and coordinate appropriately. After some generations (around 30) the system’s behaviour becomes stable. No transitions are observed, average payoffs settle very close (2.85) to the Pareto optimal and the average  $CEM_T$  is very close to one (equals 0.9). This indicates that the learning mechanism has no problems finding CE strategies. If agents don’t learn CE is due to the strategic environment, not due to something inherent to the implementation of the GA. Fig. D shows graphically this information (only 200 generations are reported due to the model becoming very stable and not presenting relevant changes).

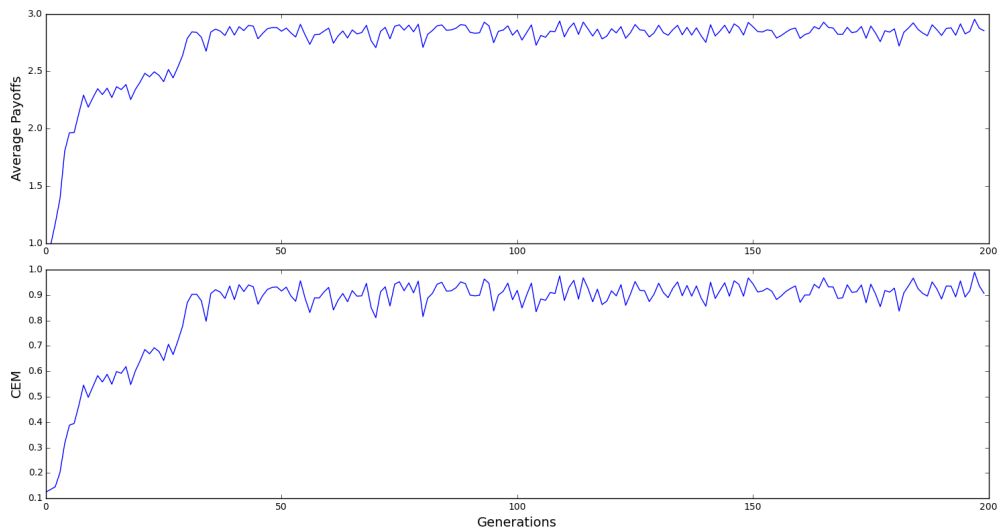
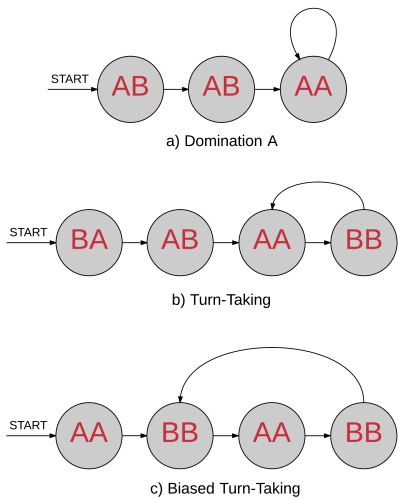


Fig. D: Average payoffs and CEM for the learning test. Around generation 30 agents have learned to follow the signal almost perfectly, shown by the high payoffs and high CEM.

## 7.5 Additional Joint Automata (JM)

### 7.5.1 No-Signal JMs

Compared to the JMs with the signal, the ones without are much simpler due to their deterministic nature. Fig. E presents three typical JMs that evolved, one under each regime. All JMs without the signal have a very similar “lollipop” shape: they visit several states in order, and at their end (since the automata are finite) they transition back to one that was previously visited. This last transition marks the beginning of a cycling behavior, meaning that the machine will forever repeat its actions. In our analysis, usually the JM takes one or more states that can include some miscoordination, but then enters the cycle and coordinates in a way reflected by the corresponding regime.



*Fig. E: Joint Automata under No-Signal treatment. Transitions are deterministic. The machines will eventually come back to an already visited state, cycling forever into a subset of states. Each Joint Automata corresponds to the indicated regime.*

## REFERENCES

- ANBARCI, N., FELTOVICH, N., GÜRDAL, M.Y., (2015). “Payoff Inequity Reduces the Effectiveness of Correlated-Equilibrium Recommendations.” *Work. Pap. Monash Univ.*,.
- ANDREONI, J., MILLER, J.H., (1995). “Auctions with Artificial Adaptive Agents.” *Games Econ. Behav.*, Vol. 10, pp. 39–64.
- ARIFOVIC, J., (1994). “Genetic algorithm learning and the cobweb model.” *J. Econ. Dyn. Control*, Vol. 18, pp. 3–28.
- ARIFOVIC, J., BOITNOTT, J.F., DUFFY, J., (2015). “Learning Correlated Equilibria: An Evolutionary Approach.” *Work. Pap. Simon Fraser Univ.*,.
- AUMANN, R.J., (1974). “Subjectivity and correlation in randomized strategies.” *J. Math. Econ.*, Vol. 1, pp. 67–96.
- AUMANN, R.J., (1987). “Correlated Equilibrium as an Expression of Bayesian Rationality.” *Econometrica*, Vol. 55, pp. 1–18.
- AXELROD, R., (1980). “Effective Choice in the Prisoner’s Dilemma.” *J. Conflict Resolut.*, Vol. 24, pp. 3–25.
- AXELROD, R., (1986). “An Evolutionary Approach to Norms.” *Am. Polit. Sci. Rev.*, Vol. 80, pp. 1095–1111.
- BONE, J., DROUVELIS, M., RAY, I., (2013). “Co-ordination in 2 x 2 Games by Following Recommendations from Correlated Equilibria.” *Work. Pap. 12-04R, Univ. Birmingham*,.
- BROWNING, L., COLMAN, A.M., (2004). “Evolution of coordinated alternating reciprocity in repeated dyadic games.” *J. Theor. Biol.*, Vol. 229, pp. 549–57.
- CAMERER, C.F., (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.
- CASON, T.N., SHARMA, T., (2007). “Recommended play and correlated equilibria: an experimental study.” *Econ. Theory*, Vol. 33, pp. 11–27.
- DUFFY, J., (2006). “Chapter 19 Agent-Based Models and Human Subject Experiments,” in: Tesfatsion, L., Judd, K.L. (Eds.), *Handbook of Computational Economics*. Elsevier, pp. 949–1011.
- DUFFY, J., FELTOVICH, N., (2010). “Correlated Equilibria, Good and Bad: An Experimental Study.” *Int. Econ. Rev. (Philadelphia)*, Vol. 51, pp. 701–721.
- DUFFY, J., LAI, E.K., LIM, W., (2014). “Language and Coordination: An Experimental Study.” *Work. Pap. Univ. Pittsbg.*,.
- FOSTER, D.P., VOHRA, R. V., (1997). “Calibrated Learning and Correlated Equilibrium.” *Games Econ. Behav.*, Vol. 21, pp. 40–55.
- FUDENBERG, D., LEVINE, D.K., (1999). “Conditional Universal Consistency.” *Games Econ. Behav.*, Vol. 29, pp. 104–130.
- GIGERENZER, G., HERTWIG, R., PACHUR, T., (2011). “Heuristics: The foundations of adaptive behavior.”
- GIGERENZER, G., TODD, P., GROUP, A.B.C.R., (2002). *Simple Heuristics That Make Us Smart*. Oxford University Press.

- GODE, D.K., SUNDER, S., (1993). "Allocative Efficiency of Markets with Zero-Intelligence Traders: Market as a Partial Substitute for Individual Rationality." *J. Polit. Econ.*, Vol. 101, pp. 119–137.
- HANAKI, N., (2006). "Individual and Social Learning." *Comput. Econ.*, Vol. 26, pp. 31–50.
- HARRISON, M.A., (1965). "Introduction to switching and automata theory." McGraw-Hill, New York.
- HART, S., MAS-COLELL, A., (2000). "A Simple Adaptive Procedure Leading to Correlated Equilibrium." *Econometrica*, Vol. 68, pp. 1127–1150.
- HOLLAND, J.H., (1975). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence.* U Michigan Press.
- HOLLAND, J.H., (1992). "Genetic Algorithms." *Sci. Am.*, Vol. 267, pp. 66–72.
- HOLLAND, J.H., HOLYOAK, K.J., NISBETT, R.E., THAGARD, P.R., (1986). *Induction: processes of inference, learning, and discovery.* MIT press, Cambridge, MA.
- HOLLAND, J.H., MILLER, J.H., (1991). "Artificial Adaptive Agents in Economic Theory." *Am. Econ. Rev.*, Vol. 81, pp. 365–370.
- IOANNOU, C.A., (2013). "Coevolution of finite automata with errors." *J. Evol. Econ.*, Vol. 24, pp. 541–571.
- IOANNOU, C.A., ROMERO, J., (2014a). "A generalized approach to belief learning in repeated games." *Games Econ. Behav.*, Vol. 87, pp. 178–203.
- IOANNOU, C.A., ROMERO, J., (2014b). "Learning with repeated-game strategies." *Front. Neurosci.*, Vol. 8, pp. 212.
- KANDORI, M., MAILATH, G.J., ROB, R., (1993). "Learning, Mutation, and Long Run Equilibria in Games." *Econometrica*, Vol. 61, pp. 29–56.
- KIRMAN, A., VRIEND, N., (2000). "Learning to Be Loyal. A Study of the Marseille Fish Market," in: Gatti, D., Gallegati, M., Kirman, A. (Eds.), *Interaction and Market Structure.* Springer Berlin Heidelberg, pp. 33–56.
- KOLLMAN, K., MILLER, J.H., PAGE, S.E., (1992). "Adaptive Parties in Spatial Elections." *Am. Polit. Sci. Rev.*, Vol. 86, pp. 929–937.
- KOLLMAN, K., MILLER, J.H., PAGE, S.E., (1997). "Political Institutions and Sorting in a Tiebout Model." *Am. Econ. Rev.*, Vol. 87, pp. 977–992.
- LAU, S.-H.P., MUI, V.-L., (2008). "Using Turn Taking to Mitigate Coordination and Conflict Problems in the Repeated Battle of the Sexes Game." *Theory Decis.*, Vol. 65, pp. 153–183.
- LEYTON-BROWN, K., SHOHAM, Y., (2008). *Essentials of game theory: Synthesis Lectures on Artificial Intelligence and Machine Learning*, 1st editio. ed, Synthesis Lectures on Artificial Intelligence and Machine Learning (book 3). Morgan and Claypool Publishers.
- MILLER, J.H., (1988). "The Evolution of Automata in the Repeated Prisoner's Dilemma." *Two Essays Econ. Imperfect Information, Ph.D. University Michigan.*
- MILLER, J.H., (1996). "The coevolution of automata in the repeated Prisoner's Dilemma." *J. Econ. Behav. Organ.*, Vol. 29, pp. 87–112.



- MILLER, J.H., BUTTS, C.T., RODE, D., (2002). "Communication and cooperation." *J. Econ. Behav. Organ.*, Vol. 47, pp. 179–195.
- MILLER, J.H., MOSER, S., (2004). "Communication and coordination." *Complexity*, Vol. 9, pp. 31–40.
- MITCHELL, M., (1998). An introduction to genetic algorithms. MIT press.
- MOORE, E.F., (1956). "Gedanken-experiments on sequential machines." *Autom. Stud.*, Vol. 34, pp. 129–153.
- MYERSON, R.B., (1997). *Game Theory: Analysis of Conflict* (1st paperback edition). Harvard University Press.
- POTEETE, A.R., JANSSEN, M.A., OSTROM, E., (2010). *Working Together: Collective Action, the Commons, and Multiple Methods in Practice*. Princeton University Press.
- RUBINSTEIN, A., (1986). "Finite automata play the repeated prisoner's dilemma." *J. Econ. Theory*, Vol. 39, pp. 83–96.
- SOLAN, E., VOHRA, R. V., (2002). "Correlated equilibrium payoffs and public signalling in absorbing games." *Int. J. Game Theory*, Vol. 31, pp. 91–121.
- SOLÉ, R., (2016). "Synthetic transitions: towards a new synthesis." *Philos. Trans. R. Soc. London B Biol. Sci.*, Vol. 371.
- SORNETTE, D., (2004). *Why Stock Markets Crash: Critical Events in Complex Financial Systems*. Princeton University Press.
- ZHANG, W., (2015). "Can Errors Make People More Cooperative? Cooperation in an Uncertain World." *Work. Pap. Purdue Univ.*,.