



Social aspiration reinforcement learning in Cournot games

Enrique Fatas¹ · Antonio J. Morales² · Ainhoa Jaramillo-Gutiérrez³

Received: 14 September 2022 / Accepted: 31 January 2024 / Published online: 8 May 2024
© Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

We offer theoretical and experimental evidence showing that *social aspiration learning* converges to the Walrasian outcome in Cournot games. Reinforcement learning converges to the competitive outcome because the Walrasian quantity is the only action that never yields profits below the average profits in the market. Using data from laboratory experiments, we show that when provided with information about average market profits, they positively (negatively) reinforce actions that yield payoffs above (below) the average payoffs in the market. When provided with both rivals' individual performance and average market profits, both heuristics (imitation and social learning) are combined by subjects and prices and profits are driven further into competitive levels, closer to the Walrasian quantity. Subjects' tendency to adjust their choices following the social learning heuristics survives and they adopt it as frequently as imitation when both predictions collide.

Keywords Experiments · Cournot competition · Walrasian convergence · Reinforcement learning · Endogenous aspiration level · Social comparison

JEL Classification C9 · L13

✉ Enrique Fatas
enrique.fatas@universidadeuropea.es

Antonio J. Morales
amorales@uma.es

Ainhoa Jaramillo-Gutiérrez
jaramill@uji.es

¹ Behavioral Economics Institute, Universidad Europea de Valencia, Valencia, Spain

² Department of Economics, University of Malaga, Málaga, Spain

³ Department of Economics and LEE, University Jaume I of Castellón, Castellón, Spain

1 Introduction

Relative to Bertrand competition, Cournot games are useful for understanding dynamic economic choices because they empirically allow for more competitive and more collusive behavior than Nash. A variety of learning rules predict departures from the Nash prediction. Vega-Redondo (1997) shows that imitation of the most successful firm in the market leads to the *most competitive outcome*, away from the Nash prediction.¹ A second set of relevant heuristics is *reinforcement learning with aspiration levels*, (Bush and Mosteller 1955) where players positively (negatively) reinforce choices that pay off above (below) a certain aspiration level (Roth and Erev 1995; Erev and Roth 1998; Karandikar et al. 1998; Börgers and Sarin 2000). When the population average payoff across markets is used as aspiration level (as in Palomino and Vega-Redondo 1999), Oechssler (2002), building on Dixon (2000), play converges to *full collusion*, away from the Nash prediction and in the opposite direction from the competitive outcome, predicted by imitate-the-best models.

In this paper, we extend this literature analyzing reinforcement learning with aspiration levels in Cournot markets when firms use a different aspiration reference point: the average market payoff. For a variety of reinforcement learning rules, we show that when the aspiration level is set to the average market payoff, firm behavior becomes more competitive rather than more collusive. This result follows a property of the Walrasian outcome described in Morales and Fernandez-de-Cordoba (2012): as a firm choosing the Walrasian outcome will never get a payoff below the average market payoff, the Walrasian outcome will never be negatively reinforced. Conversely, collusive outcomes will never be positively reinforced, as they never yield profits above the average market payoff. We prove that markets converge to the Walrasian outcome for two models of learning: “Win-Stay, Lose-Switch” (WS-LS hereafter), originally proposed by Thorndike (1911) and studied in Oechssler (2002) and the reinforcement learning rule proposed by Roth and Erev (1995) (RE hereafter).²

We illustrate how these theoretical properties apply to short run settings by simulating competition in Cournot markets. We use the 3-player symmetric Cournot game in Apesteguia et al. (2007), with no production cost and five available actions and simulate 100 markets over 60 rounds. We study WS-LS and RE learning rules in three different scenarios: with no aspiration level, and with aspiration levels set by average market and average industry payoffs (being an industry composed of nine firms in three markets).

With no aspiration level, RE makes firms to stay around the Nash equilibrium.³ When aspirations are set by average industry payoffs, firms do not converge towards collusive prices in the short run, but rather linger around the Nash equilibrium price

¹ Reviewing the imitation literature is far beyond the objectives of this paper. Alós-Ferrer and Shi (2012) gives a brief overview of the seminal papers.

² Note that in this paper we do not derive a full rational model of imitation. Hedlund and Oyarzun (2018) is a good example of a bounded rationality approach to imitation dynamics, with an heterogeneous population of players and dual incomplete learning.

³ No aspiration level means that the played action is always positively reinforced (i.e. this is equivalent to setting the aspiration level to 0 or lower than the minimum payoff from the game). The dynamics for the WS-LS rule with no aspiration level is trivial, as players will stick to the action played in the first round.

(for both the WS-LS and the RE rules). When aspirations are set by the average market profit, prices steadily tend towards the Walrasian outcome, away from Nash equilibrium, confirming the short run relevance of the long-term theoretical predictions.

We test these results using the experimental setting in Apesteguia et al. (2007) in the lab. Our design features five treatments. In the baseline, *No info*, participants only receive information about their own payoffs and the vector of actions chosen in the market. In *Market (Industry)*, participants receive additional information: the average market profit (the average industry profit), exogenously setting a social aspiration reference point. In *Selection*, participants receive the same information as in *Market*, but the worst performer is kicked out of the market and remains inactive for one period (earning zero profits, without receiving any information). Finally, in *Full Info*, participants are informed about the average market payoff and about the individual payoffs received by other participants in her market, opening the door for imitation of the best.

While *Market* and *Industry* set aspirational reference points, *Selection* makes the consequences of competition more dramatic, increasing any competitive effect of social comparisons. As in *Full Info* imitation of the most successful firm is possible, we disentangle between imitation (if the action of the most successful firm is copied) and social learning (if participants follow one of the WS-LS and RE rules). In this treatment, there are instances where participants may learn the best performing action and therefore “imitate” it or adopt social learning and follow either WS-LS or RE rules. In a subset of these instances, both models predict different actions for the next round, so we can neatly disentangle whether one of the models predict the actions of participants better than the other.

Our experimental results neatly confirm our simulations, showing that firms easily compare themselves with rivals assessing the relative market performance. Without information about market profits (as in *No info*), market prices stay around the Nash-Cournot equilibrium, and the modal action is the Cournot quantity. With information on the average market profit (as in *Market*), participants choose larger quantities and prices and profits tend towards competitive levels (and the modal action is the Walrasian quantity). Moreover, when subjects receive information about the average industry profit, market prices stay around the Nash-Cournot payoffs, and the modal action is the Cournot quantity. The levels of imitation in *Full Info* and in *Market* are remarkably similar, and in our analysis of imitation and social learning heuristics we conclude that even in cases in which firms may uniquely identify the most successful action in the previous round, social learning is as powerful as imitative heuristics, surviving the introduction of individual profits.

The observed effect of social competition in our experiments extends the importance of relative performance to market settings. Psychologists and management scholars have long recognized that individuals (say, workers) are not exclusively motivated by absolute monetary rewards (Etzioni 1971; Deci 1975). Bicchieri et al. (2021) review how individuals compare their actions with the expected actions of others and adjust their behavior to these expectations. The evidence about the interaction between relative performance and intrinsic motivation in other domains is overwhelming. A taste for higher rank may be a powerful behavioral force even when agents are paid according to absolute rather than relative performance. Individuals may exhibit loss aversion

in social comparisons: negative deviations from a reference point count more than positive deviations.⁴

Therefore, in this paper, we contribute to the more general experimental literature on social comparisons.⁵ A large branch of this literature shows that the revelation of information about the behavior of other players may trigger non-trivial dynamics; the selection of one dynamic is sensitive to the task type and to the social information disclosed.⁶ In portfolio decisions, Dijk et al. (2014) find that information about relative performance makes over-performers prefer less risky portfolios and underperformers to be more prone to choosing riskier ones. In public good settings, the availability of information on individual contributions reinforces conditional cooperation, a well-documented behavioral pattern. Among others, Andreoni and Petri (2004), Croson et al. (2005, 2015) and Fischbacher and Gächter (2010) show that this type of social information promotes imperfect convergence in cooperation levels.⁷

Experimental research on the impact of social comparisons in market settings is surprisingly scarce. Schoenberg and Haruvy (2012) show that relative performance information had a significant impact on the outcome of a double auction asset market: prices were higher in markets in which traders observed the best performer as compared to those in which traders received information about the performance of the worst. Gortner and van der Weele (2019) and Lindner et al. (2021) analyze social comparisons in experimental financial markets to find that the presence and framing of social information is an important determinant of equilibrium in financial markets and investment decisions, respectively. Haruvy (2019) shows that relative payoffs play a role in two-sided matching markets. Our paper adds to this literature by showing that providing information of relative performance has a significant impact on the outcome of the Cournot market: competition is fiercer and consequently prices are lower, without imitation of the most successful rival or rivals. Finally, Hedlund (2015) shows how the market outcome becomes less competitive as the number of markets from which

⁴ Loss aversion is one of the foundations of Prospect Theory and it has been documented as a powerful driver in many economic/psychological situations (see the seminal contribution by Kahneman and Tversky 1979). Buunk and Gibbons (2007) surveys the extensive research on the relevance of social comparison. Social comparisons also lie at the heart of Inequity Aversion in Fehr and Schmidt (1999). Fallucchi et al. (2021) shows how inequity aversion may be asymmetric in experimental games (stronger when breaking a group rule).

⁵ There is evidence that in situations where subjects compare to inferior performers, they are more satisfied with performance as when they compare to superior performers, see Blanton et al. (1999) and Buunk and Gibbons (2007). The social comparison heuristic implies that subjects are more satisfied and reinforced the current behaviour with their performance when they earn more than the average, and less satisfied and less eager to reinforce when they are below the average.

⁶ Bilancini and Boncinelli (2018, 2020) and Fatas et al. (2020) study the emergence of conventions in social dilemma games. There are also field studies. For example, Tran and Zeckhauser (2012) and Azmat and Iriberry (2010) reported field experiments using high school students as subjects, whereas Blanes and Nossol (2011) focused on workers' performance in firms, and Restrepo-Plaza and Fatas (2022) study cooperation among victims of conflict in Colombia.

⁷ However, not all the experimental evidence reports a substantial and positive effect. Erikson et al. (2009) found that relative performance feedback on employee effort has no effect. Charness et al. (2013) documented the dark side of competition for status in an experiment where workers could sabotage others, or artificially inflate their own output. Fatas et al. (2018) document how the source of information may push individuals towards conformity (with the observed action) or differentiation (from it).

firms can imitate increases (in other words, as firms receive more information about firms in other Cournot oligopolies).

The rest of the paper is as follows. In Sect. 2 we present the theoretical analysis of reinforcement learning with social aspiration levels and show that when the market average profit plays the role of an aspiration level, it affects the convergence of the stochastic process. Experimental results are presented in Sect. 3. Section 4 analyses the existence of imitation in our experiment and finally Sect. 5 concludes.

2 Convergence theorems and simulations

We consider a standard symmetric Cournot market in which n firms with cost function $c(q)$ face a downward-sloping demand function $p(Q)$. There is a unique symmetric Walrasian quantity q^w , defined as the quantity at which price equals marginal costs when all firms produce the same quantity, $p(nq^w) = c'(q^w)$. Firms repeatedly play the Cournot game in discrete time. For technical reasons, we discretize the action space and assume that the largest quantity that a firm can select is the Walrasian quantity $\{0, \dots, q^w\}$.⁸

2.1 Pure strategy reinforcement learning with social aspiration level and stochastic stability of the Walrasian state

Vega-Redondo (1997) was the first to prove that in the Cournot game, an imitation dynamic would drive the market to the Walrasian outcome.⁹ He did so by using a pure strategy learning rule where each firm enters a period with a pure strategy. After observing the actions and profits to each participant in the market, firms enter next period with the pure strategy that achieved the highest payoff in the current period.

In our setup we assume that after each period $t = 1, 2, \dots$ firms observe their own payoff and the market average profits. Hence, firms cannot use an imitation dynamic as full information about the output and profits of all participants is not available. There are however models that incorporate aggregate information. Thorndike (1911) propose the win-stay, lose-switch rule (WS-LS hereafter) where a decision maker chooses next period the same action if the payoff received in the current period is not smaller than an aspiration level, and changes to another action otherwise. Oechssler (2002), building on Dixon (2000), study the convergence properties of the WS-LS rule, where the aspiration level is the average industry profit (where an industry is composed of several markets). In this paper, we use the WS-LS rule but set the aspiration level to the market average profit where firms are operating.

Each firm enters the current period with a pure strategy, which is the output that she will produce. If the current payoff to a firm is equal or larger than the market

⁸ This assumption simplifies the proofs because the market-clearing price will never be below the marginal cost. This assumption is experimentally sound as firms will never get negative profits. This is why it is imposed in the experimental design in Apesteguia et al. (2007, 2010). We also assume it in our experimental setting and simulations, which replicate the experimental design in Apesteguia et al. (2007).

⁹ Alos-Ferrer and Ania (2005) is a canonical analysis of the evolutionary and dynamic foundations for competitive behavior in symmetric market games.

average profit, the firm chooses the same quantity next period, otherwise the firm experiments randomly with an alternative action. To use techniques from stochastic stability (Samuleson, 1997), it is further assumed that with probability $\varepsilon > 0$ each firm forgets the action prescribed by the learning rule and chooses a quantity at random.

For $\varepsilon > 0$, the dynamic process follows a irreducible and aperiodic Markov chain which has a unique limit distribution. The following proposition shows that when the error probability ε goes to 0, the limit distribution puts mass 1 on the state in which all firms choose the Walrasian output.

Proposition 1 *The Walrasian state is the unique stochastically stable state when the market average profit is the aspiration level in the WS-LS rule.*

Proof Only absorbing states of the unperturbed dynamics are candidates to be stochastically stable. Under the reinforcement learning rule with social aspiration level, a state is absorbing if and only if all firms are receiving the same payoff. We use the mutation counting technique in Freidlin and Wentzell (1984). Consider any absorbing state $q^a = (q_1, q_2, \dots, q_n)$ different from the monomorphic state where all firms choose the Walrasian quantity q^w . Let p^a denote the price associated to the absorbing state. We have for all firms i and j ,

$$p^a q_i - c(q_i) = p^a q_j - c(q_j)$$

Recall that we are assuming that quantities are bounded by the Walrasian quantity. Without loss of generality, assume that quantities in the absorbing state q^a are in increasing order: $q_1 \leq q_2 \leq \dots \leq q_n \leq q^w$ with at least one strict inequality. Let K^- denote the set of firms choosing a quantity strictly smaller than the Walrasian quantity q^w . We will show that a single mutation is enough to destabilize this absorbing state and put it in the way to the Walrasian state. To see this, consider one mutation from a firm belonging to K^- , that increases its production to the Walrasian quantity. The new price p^m will be smaller than p^a and in the new state, the firm belonging to K^- with the smallest quantity will be the firm earning the lowest profits -and therefore, earning less than the average profit in the market-. With positive probability this firm will choose in the next period the Walrasian quantity, yielding a lower price $p^m < p^a < p^M$. With this price, the firm earning the lowest profit in the market will be the firm belonging to the set K^- with the lowest quantity... and so on. This process will eventually lead to all firms choosing the Walrasian quantity.¹⁰ Consider next the Walrasian state in which all firms choose the Walrasian quantity q^w . Note that if there is only one mutation, the

¹⁰ The assumption that players cannot choose quantities above the Walrasian quantity implies that the price will never be below the marginal cost and therefore that in the sequence triggered by the initial mutation, the firm with the lowest earnings is the firm in K^- with the lowest quantity. If we dropped this assumption, then we would need to define the set K^+ of firms choosing quantities larger than the Walrasian quantity in the absorbing state and allow for the possibility that in the dynamic sequence triggered by the initial mutation, the price might fall below the marginal cost. In the event of negative profits, then firm with the lowest earnings will be the firm in K^+ with the largest quantity, which with positive probability will choose the Walrasian quantity in the next period. This is how with positive probability, all firms in $K^- \cup K^+$ will end up choosing the Walrasian quantity.

system will return to the Walrasian state, as the mutant firm will get below the average profit in the market. $\square$

Three things to note about this result. First, it nicely complements the result in Vega-Redondo (1997) by showing that imitation is not needed for the stochastic process to spend most of the time in the Walrasian state when the probability of mistakes ε is small.

Second, our result contrasts with Oechssler (2002), which finds that the stochastic process yields collusion rather than competition. Why this stark discrepancy when both analyses are based on the same reinforcement learning rule? To see it, consider two markets with n firms and the monomorphic state with all firms in all markets choose the Walrasian quantity. All firms in all markets earn 0 and the average industry profit is 0. It is easy to check that one mutation in one market will suffice to drive the whole industry to the collusive outcome. A mutant firm choosing any quantity below the Walrasian quantity will make all firms in her market (name it the *perturbed* market) to earn positive profits. Then, all firms in the “unperturbed” markets will earn less than the average industry profit and, with positive probability, they will choose the collusive outcome. Next, all firms in the perturbed market will earn less than the average industry payoff and will switch with positive probability to the collusive quantity.¹¹

Third, the simple pure strategy reinforcement WS-LS rule with social aspiration level implies a basic dynamic where the switching probabilities are independent from the magnitude of the dissatisfaction (defined as the distance from the current payoff to market average payoff). In addition, no firm should switch away from actions that yielded payoffs above the average market payoffs, unless it is done by mistake.

We next show that we can find other reinforcement learning rules with larger state spaces that also converge to the Walrasian quantity and that display a more sophisticated dynamic. In particular, the probability of switching away from the chosen action will depend positively on how poorly the action performed compared to the market average. In addition, for positively reinforced actions, there will always be a positive probability of the firm switching away from it in the next period (as long as the state of the learning rules is not a degenerated mixed strategy).

2.2 More general reinforcement learning with social aspiration level and almost sure convergence to the Walrasian state

Reinforcement learning was first formulated by psychologists, Bush and Mosteller (1955) and introduced in economics by Roth and Erev (1995) and Börgers and Sarin (1997, 2000). We aim at enlarging the state space of the reinforcement learning rules while maintaining the convergence result to the Walrasian outcome. We will consider the rule proposed by Roth and Erev (1995).¹² At each stage $= 1, 2, \dots$, a player must choose one action from a set of m actions. To each action and stage, it is associated a

¹¹ It is easy to show that a monomorphic state where all firms choose the collusive outcome cannot be destabilize by a single mutation. The reason is that in the “unperturbed” market, firms are getting the highest possible profit and therefore under the WS-LS rule, they will never switch away from the collusive quantity.

¹² In Appendix D, we will show a theoretical analysis and simulations results for a generalization of the rule in Börgers and Sarin (2000) with an aspiration level equal to the average market payoff. A theoretical

propensity; let $A_s(t)$ denote the propensity of action s at stage t . There is also a set of initial propensities $A_1(1), A_2(1), \dots, A_m(1)$. The probability that action s is chosen at stage t is

$$p_s(t) = \frac{A_s(t)}{A_1(t) + A_2(t) + \dots + A_m(t)}$$

After playing an action and receiving a payoff, the propensity of the chosen action is increased following a reinforcement function that in the basic model is simply the payoff received π . The updating rule for the propensities is as follows

$$A_s(t+1) = \begin{cases} A_s(t) + \pi & \text{if action } s \text{ was chosen at stage } t+1 \\ A_s(t) & \text{if action } s \text{ was not chosen at stage } t+1 \end{cases}$$

This basic rule can be described by a Polya urn. The initial propensities are the initial numbers of colored balls in the urn, where different color corresponds to different actions. At each stage, a ball is randomly extracted from the urn and a number of balls equal to the payoff received of the same color as the chosen one are added to the urn.

This class of reinforcement learning rules can be generalized to accommodate aspiration levels. In this paper, we set an aspiration level equal to the average market profit. This implies that the propensity of the chosen action by player i will be reinforced by a function $R_i(\pi_1, \dots, \pi_n) = \pi_i - \frac{\pi_1 + \dots + \pi_n}{n}$, where π_i is the payoff received by player $i = 1, 2, \dots, n$. In terms of the Polya urn, this implies that some balls are extracted from the urn if the received payoff falls short of the aspiration level.

In general, the analysis of learning rules described by Polya rules is complex because one needs to keep track of several urns, one for each player (see Pemantle 2007 for a literature review).¹³ However, in our case, where the aspiration level is set to the market average payoff, the analysis is easier, as we can build on work by Beggs (2005). Consider the parametrization in Apesteguia et al. (2007), with costless production and linear demand $P = 120 - Q$. For a triopoly, the collusive quantity is 20, the Nash quantity is 30 and the Walrasian quantity is 40. The profit to firm i is $\pi_i(q_i, q_{-i}) = (120 - Q)q_i$

Footnote 12 continued

analysis for this rule is more complex. In the literature, there are only a few theoretical results, basically for 2×2 games. When the state space is the simplex, Börgers and Sarin (1997) study the convergence properties of the rule in (Cross 1973), a particular member of the reinforcement learning rules proposed by Bush and Mosteller (1955), where all actions are always positively reinforced. Izquierdo et al. (2007) study convergence of reinforcement learning in 2×2 games with fixed aspiration level and Börgers and Sarin (2000) study convergence with endogenous aspiration levels for decision problems with two actions. Duersch et al. (2010) document how imitation is the only exploitation free learning algorithm when subjects interact with different learning algorithms (best response, fictitious play, imitation, reinforcement learning) in a large online experiment where subjects can influence algorithms to their advantage.

¹³ There are generalizations of the reinforcement rule that incorporate for example recency (Erev and Roth, 1995, 1998), forgetting (Roth and Erev 1995; Erev and Roth 1998; Rustichini 1999), inertia (Bender et al. 2001a, b) as well as aspiration levels (Bender et al. 2001a, b). See for example Erev and Haruvy (2013) for a review. In this literature, even when the rules are applied to strategic settings, the aspiration level is either fixed (Izquierdo et al. 2007) or typically depends on individual payoff histories (see Börgers and Sarin 2000, Posch et al. 1999 or Karandikar et al. 1998). Bendor et al. (2001a) surveys the literature on reinforcement learning with aspiration levels.

and the reinforcement function, i.e. the number of balls added/extracted for each action profile can be written in terms of actions as follow

$$R_i(q_1, q_2, q_3) = \frac{1}{3}(120 - Q)(3q_i - Q)$$

We will show that there exists a relation $\succ_{SD}$ of strict dominance defined over quantities, with the Walrasian quantity strictly dominating all other quantities in terms of reinforcement values. To see it, we compute the derivative of the reinforcement function with respect to q_i and show that it is positive for any profile of quantities, implying that the maximum of the reinforcement function is achieved at the Walrasian quantity. This derivative is

$$\frac{\partial R_i}{\partial q_i} = \frac{1}{3}(120 - Q)(120 - 3q_i)$$

which is positive because the individual quantities are bounded above by 40, the Walrasian quantity.

Intuition suggests that the Walrasian quantity will be reinforced more than any other action and that this will lead the system to converge to the probability of choosing the Walrasian quantity to 1. We can use Theorem 3 in Beggs (2005) to prove this result for the Cournot setting in Apesteguia et al. (2007).¹⁴

Proposition 2 *In the Cournot setting in Apesteguia et al. (2007), under the RE rule with social aspiration level, for each player, the probability of playing the Walrasian quantity converges to 1.*

Proof Theorem 1 in Beggs (2005) focuses on decision problems with positive payoffs, and consider the RE rule without aspiration level, in which actions are reinforced by the payoff received. For this case, Beggs (2005) proves that if action 1 gets in expected terms, larger payoffs than action 2, i.e. if one action strictly dominates the other action, then the propensity of action 1 grows at a faster rate and the probability of action 1 relative to the probability of action 2 converges to 1 (using a super-martingale theorem). Theorem 2 in Beggs (2005) extends this result to any number of actions. And Theorem 3 shows that RE rule will learn to play the dominant strategy in the long run, i.e., the probability and empirical frequency with which a player plays a strictly dominant action converges to 1. In our case, the update in propensities is given by the reinforcement function matrix and there is one action, the Walrasian quantity, that strictly dominates all others. So, Theorem 3 applies. $\square$

In the next section, we investigate the short-run properties of these reinforcement learning rules using simulations in triopolies with five actions.

¹⁴ The above argument for the strict dominance of the Walrasian quantity in terms of the reinforcement function can be easily extended to linear Cournot games. Our intuition is that it also extends to general Cournot games. Because the goal of this paper is to show the experimental relevance of social learning reinforcement, we leave the theoretical analysis of the general case for future research.

2.3 Simulations

The simulation exercise is pertinent because the theoretical results refer to limiting cases (i.e. vanishing noise for the WS-LS rule and for time going to infinity for the RE rule), and it is not clear that these limiting cases are relevant to explain experimental data.

We use the symmetric Cournot game in Apesteguia et al. (2007) with three firms, five actions (a, b, c, d , and e), linear demand linear demand function $p(Q) = 120 - Q$ and zero marginal cost. The five actions are $a = 20$ (Collusive quantity), $b = 23$, $c = 30$ (Cournot-Nash quantity), $d = 36$ and $e = 40$ (Walrasian quantity). For this parametrization, the Collusive price is 60 (action a), Cournot-Nash price is 30 (action c), and the Walrasian price is 0 (action e).

For *No Info* and *Market*, we simulate 100 markets of three firms each over 60 periods. We focus on the evolution of the market price over 60 rounds, averaged over 100 markets, as well as on the distribution of actions for the second part of the simulation (periods 31–60). For *Industry*, as one industry is composed of 3 markets with 3 firms each, we simulate 33 industries (99 markets). For all learning rules considered, the initial conditions are chosen randomly. We will be more specific about the randomization process of initial conditions when discussing the simulation results for each rule.

We start with the simulations for the WS-LS rule. This rule has two parameters: the initial action and the noise parameter. Figure 1 displays the simulation results when the initial action is chosen randomly from a uniform distribution over the five actions and the noise parameter is set to 0.2. This noise level implies that with probability 80%,



Fig. 1 Simulations from the “Win-Stay, Lose-Switch” learning rule. Left panel plots market prices obtained in the simulation of the “Win-Stay, Lose-Switch” learning rule. Right panel displays the relative frequencies of quantities in rounds 31–60 for the Market info (blue stack) and Industry (dark pink stack) treatments

a firm will play the quantity prescribed by the rule, and that with the complementary probability, 20%, the firm will choose a random quantity.

Because initial conditions are chosen randomly, average market prices start around the midpoint of the price interval $[0, 60]$ in period 1. When the aspiration level is the average market payoff, simulations show a quick drop in the market price over the first few rounds -as predicted by Proposition 1- and then the market price stabilizes. For the Industry condition, the average market price oscillates around the Nash equilibrium price with no tendency towards collusive prices, showing that the theoretical prediction in Oechssler (2002) is not observed in our setting and for our choice of parameters.

The right panel in Fig. 1 displays the relative frequencies of quantities in rounds 31–60. Simulations for *Market* show that the distribution is skewed to the right, with the modal action being the Walrasian quantity (action *e*), with a proportion above 50%, well above the other quantities (the second in importance is quantity *d*, with a frequency short of 20%). In *Industry*, we observe a flat symmetric distribution, and the (weak) modal action is the Cournot quantity (action *c*).

Appendix A contains a sensitivity analysis of the simulations to changes in the noise parameter. We consider seven different values: 0.4, 0.2, 0.1, 0.05, 0.02, 0.01 and 0.005. In *Market*, prices always drop over the first few rounds and the noise parameter sets the level at which the market price stabilizes: the lower the noise parameter, the lower that price level. In *Industry*, we only find a slightly upward trend towards collusive prices when the noise parameter is below 1%.

Figure 2 displays the simulation results for the RE rule. Together with *Market* and *Industry*, we also have simulations for the basic RE rule without aspiration levels to simulate situations in which firms have no information about rivals' performance (as

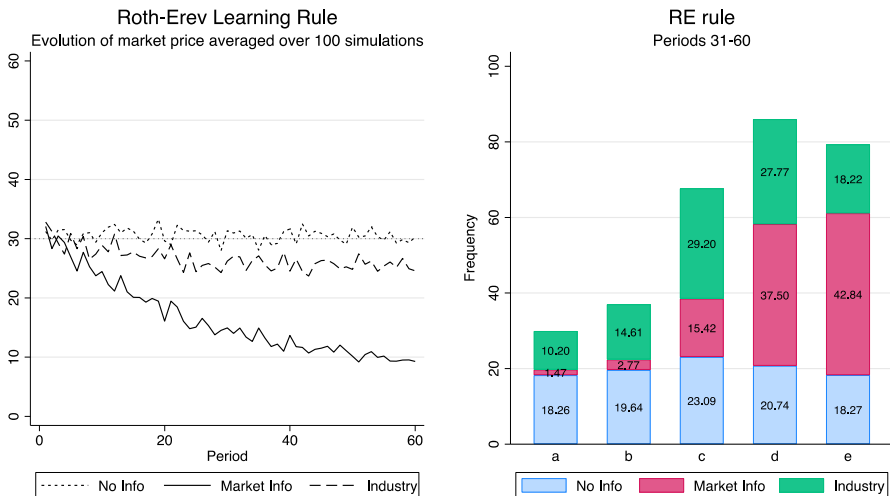


Fig. 2 Simulations from the Roth-Erev reinforcement rule. Left panel displays the market prices obtained in the simulations of the Roth-Erev reinforcement rule. Right panel displays the relative frequencies of quantities in rounds 31–60 for the No Info (light blue stack), Market info (dark pink stack) and Industry (green stack) treatments

in *No info*). The parameters of the RE rule are the set of initial propensities and the reported simulations correspond to the case in which the initial propensity of each action is a random value from the uniform interval $[0, 1600]$, where 1600 is the largest profit one firm can obtain in the market and 0 the lowest.

Figure 2 shows that all simulations start around the midpoint of the interval $[0, 60]$. Without aspiration levels (*No Info*), prices oscillate around Nash equilibrium. When the aspiration level is set to the average market profit, prices steadily drop, consistent with our theoretical results of convergence to the competitive price. Finally, when the aspiration level is set at the industry level, prices show a slight decline although they stay close to the Nash equilibrium price.

The distribution of quantities for periods 31–60 are displayed in the right panel of Fig. 2. With no aspiration level, the distribution is flat with the Cournot quantity as the modal quantity. When the aspiration level is set at the average market profit, we observe again a skewed distribution towards competitive actions: the modal action is the Walrasian quantity (42% of observations), followed closely by action *d* (37.5%). Finally, when the aspiration level is set at the average industry profit, the modal action is Cournot (29.2%), although action *d* is chosen similarly (27.8%).

In summary, our simulations show that the asymptotic results described in the theory section for the reinforcement learning rules considered are relevant for the description of the short-run behavior of the market game in two scenarios: when there is no aspiration level and when the aspiration level is set at the average market profit. No sizable trace of collusive behavior is observed when the aspiration level is set to the average industry profit. Note that simulations for both reinforcement learning rules are qualitatively similar (tendency towards more competitive prices) but are quantitatively different. The pure-strategy reinforcement learning rule WS-LS stabilizes around a flat price after a few rounds while the simulated prices for the more general RE learning rule steadily declines to the Walrasian equilibrium price $p = 0$.

3 Experimental design, procedures, and hypotheses

The experiment is composed of one game and five treatments. Table 1 contains the description of our experimental design.

3.1 Experimental design

We follow the same design as in Apesteguia et al. (2007). Subjects repeatedly play (during 60 periods) a simple 3-player game with five actions derived from a symmetric Cournot game. Each round, participants choose one quantity in a symmetric Cournot game with five actions (*a*, *b*, *c*, *d*, and *e*) corresponding to $a = 20$ (Collusive quantity), $b = 23$, $c = 30$ (Cournot-Nash quantity), $d = 36$ and $e = 40$ (Walrasian quantity). Subjects are not informed about the underlying payoff structure, but they are aware that the payoff process remains constant along the duration of the experiment. Interaction in the experiment takes place within groups or cohorts of 9 participants each (12 participants in the *Selection* treatment). At the beginning of the experiment subjects

Table 1 Experimental design

	Treatment				
	No info	Market	Full info	Selection	Industry
Subjects	45	45	45		45
Sessions	1	1	1	60	1
Markets	15	15	15	15	15
Groups	5	5	5	5	5
Matching	Strangers	Strangers	Strangers	Strangers	Strangers
Feedback	Market action profile and one's own payoff				
		+ Avg market profit	+ Avg market profit + individual actions and profits	+ Lowest profit firm leaves the market	+ Average Industry profit

are randomly assigned into 5 groups. Each period, participants within a group are randomly matched into three-firm markets to minimize repeated game effects. We refer to each group as an industry, which is composed of three markets.

We consider five experimental treatments. In all of them but *Selection* (described later) 45 participants are assigned to 5 independent groups or cohorts of 9 participants each. In all treatments, at the end of each period participants receive information about their own individual payoffs and the market actions profile (the three letters/actions chosen in the market in that period). In addition, players in each treatment are informed about:

- *No info*: No additional information is shared.
- *Market*: Average market profit.
- *Full info*: Average market profit and profits and actions of each firm in their market.
- *Industry*: Average profit in their group or cohort of 9 players.

In *Selection*, 60 participants are divided into 5 groups of 12 players each. Each period, there are 9 active players in each cohort, which are randomly matched into three-firm markets, and 3 passive players. At the end of each period t , active players are informed about the market actions profile, their own individual payoffs, plus the average profit in the market (as in *Market*). Passive players receive no information at all. The firm with the lowest profits in each market¹⁵ in period t becomes a passive player in period $t + 1$ and passive players at period t become active in period $t + 1$.

3.2 Experimental procedures

We collected 5 independent observations per treatment, each one consisting of 2700 actions (45 participants $\times$ 60 rounds), for a total of 67,500 non-independent observations. Experiments were run in the laboratory for research in experimental economics at the University of Valencia (LINEEX) using the z-tree software (Fischbacher 2007). Two hundred and forty business and economics students were recruited using a standard electronic recruitment procedure and they were allocated randomly to any of the five treatments. Participants got an average pay of \$25 for an experiment that lasted 60 min on average.

3.3 Hypotheses

We state a set of hypotheses regarding behavior in the five treatments. In *No info*, subjects were only informed about one's own payoff and the profile of actions in the market. The theoretical literature on learning in repeated games when players know only their own payoff (*uncoupledness*) and the common history of play (*perfect monitoring*) shows positive results on convergence to Nash equilibrium (see for example, Kalai and Lehrer 1993 or Hart and Mas Colell 2006). This is consistent with the results from the simulation under the RE rule without aspiration levels. Therefore, convergence to the Cournot-Nash outcome is expected to prevail.

Hypothesis 1 Behavior will converge to the Cournot-Nash prediction in *No info*.

¹⁵ Ties are randomly broken.

In *Market*, subjects received information about the average profits in the market. We expect subjects to use the average market profits as an aspiration level. Based on Propositions 1–3 and the simulations described above, we expect behavior to move away from the Cournot-Nash prediction towards more competitive outcomes.

Hypothesis 2 In *Market*, behavior will be more competitive than in *No info*, participants will choose larger quantities, and market prices will be lower.

In *Industry*, despite the theoretical work by Oechssler (2002) that social competition yields collusion under the WS-LS rule, our simulations show that there is no tendency towards collusive actions, but rather the system lingers around the Cournot-Nash outcome, with a small decline for the RE rule.

Hypothesis 3 In *Industry*, quantities will not be significantly different from quantities in *No info*.

We expect the attraction to the Walrasian quantity to be more prominent if an additional selection mechanism is introduced. The intuition is that, according to Morales and Fernandez-de-Cordoba (2012), the Walrasian quantity is the only choice that eliminates the probability of being excluded.

Hypothesis 4 In *Selection*, behavior will be more competitive, and quantities chosen by participants will be significantly above the ones observed in *Market* (with lower market prices). Quantities (prices) will not be below (above) the ones observed in *Full Info*.

In *Full info*, players may use either a social comparison rule or the imitative rule “Imitate the Best”. Both are known to steer the market towards competitive prices. Hence, we expect more competitive prices in this treatment than in the treatment *Market* due to the addition of two imitative heuristic.

Hypothesis 5 In *Full Info*, behavior will be more competitive than in *Market* and participants will choose quantities significantly larger (with lower market prices).

We summarize our Hypotheses 1 and 5 in the following ordering of market prices in the five treatments.

$$Industry = No\ info > Market > Full\ Info \geq Selection$$

This summary highlights the role played by treatments *No info* and *Selection* as lower and upper bounds.

4 Experimental results

This section presents the experimental analysis of the results. It is organized as follows. First, data are evaluated at the aggregate level. Second, we run an analysis of individual switching ratios to assess the proposed reinforcement-learning rule with endogenous aspiration level. Estimations of reinforcement learning rules are in the Appendix.

Table 2 shows average prices for all treatments, separately for first round, all rounds and first half and second half of the experiment. Our cautious statistical analysis follows

Table 5 in Appendix E, in which we estimate treatment effects by running random-effects GLS regressions with market prices as the dependent variable (one observation per market and period), and robust standard errors clustered at the triopoly level. Bonferroni-adjusted p -values used to determine significance. As an illustration that our results are not driven by the large number of rounds, we report in Table 11 in the Appendix the outcome of non-parametric tests (Mann–Whitney) using only five independent observations per treatment, one per cohort. Both analyses yield almost identical results.

Our results show no significant differences in round 1, and no significant differences between *Industry* and *No info* (except for the first block of 30 rounds). Hypothesis 1 and 3 involving point predictions are remarkably confirmed, as average market prices in the second half of the experiment are 30.65 and 31.82 for the treatments *Industry* and *No Info* respectively. These values are extremely close to the Cournot–Nash equilibrium price (30).

However, differences across the other treatments are significant across all 60 periods and in both 30-period blocks: prices are significantly lower in *Market* and *Full Info* than in *No Info*. Market prices are not significantly different in *Full Info* and in *Selection*. Interestingly, prices are significantly lower when two competitive heuristics are possible (as in *Full Info*) than when only one is at play (as in *Market*).

We summarize the aggregate results described in Table 2 in the following three results, confirming our ordering all five treatments, for all 60 periods, and in periods 31–60.

Result 1. Without information (as in No info) and with information at the industry level (as in Industry), market prices converge to the Cournot–Nash

Result 2. With information about average market profits (as in Market), we observe more competitive (and significantly lower) market prices.

Results 3. The introduction of a selection mechanism (as in Selection) or providing a second competitive heuristic, by giving full information on individual actions and profits of all participants (as in Full Info), exacerbates this competitive drift.

Table 2 Market prices by experimental treatment

Periods	Industry	No info		Market		Full info		Selection
1	33.67 (11.23)	41.67 (5.10)		38.07 (9.26)		39.33 (4.96)		36.73 (4.04)
1–60	30.14 (2.28)	32.60 (1.98)	***	25.68 (3.41)	***	19.31 (2.29)		20.11 (1.41)
1–30	29.90 (2.42)	33.38 (2.13)	**	26.88 (3.00)	***	20.48 (2.08)	***	24.13 (1.76)
31–60	30.38 (2.63)	31.82 (2.76)	***	24.49 (4.57)	**	18.14 (2.86)		16.09 (1.43)

Robust standard errors clustered at the triopoly level in parenthesis

Significance using Wald test and Bonferroni-adjusted p values after random-effects GLS regressions

At the 5% level, *At the 1% level

Figure 3 displays the relative frequencies of actions by treatment, for the whole experiment and the second half of the experiment. Simulations predicted the Cournot Nash choice to be the modal action in two treatments, *No Info* and *Industry*, while in the others, the predicted modal action should be the Walrasian quantity. These predictions are confirmed. In *No info*, the modal action is the Cournot-Nash quantity, which is selected 27.59% of the times (being the Walrasian quantity the less frequently chosen by participants, 14.96%). In *Industry*, the modal action is also the Cournot-Nash quantity, selected 24.44% of the times. The distribution of choices is, not surprisingly, consistent with Result 1 above.

In sharp contrast, in the other three conditions the modal choice is the Walrasian quantity: 32.74% of the times in *Market*, and up to 40.63% in *Selection* and 42.11%

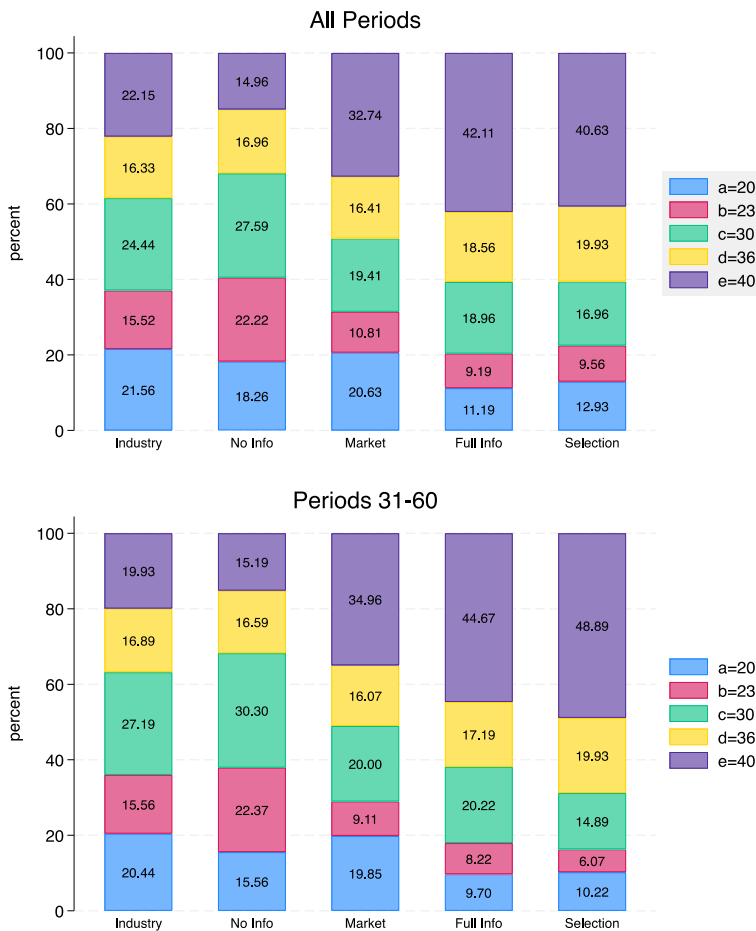


Fig. 3 Relative frequencies of actions in each experimental treatment

Table 3 Cournot-Nash and the Walrasian quantities by experimental treatment

Treatment	Average profits		Switching ratio		% below aspiration	
	Cournot	Walrasian	Cournot	Walrasian	Cournot	Walrasian
<i>No info</i>	951.30	875.45	47.54	64.39	–	–
<i>Market</i>	828.38	674.93	56.03	43.93	48.66	0
<i>Full Info</i>	703.30	496.36	59.84	36.58	65.43	0
<i>Selection</i>	726.68	525.54	58.37	35.29	61.35	0
<i>Industry</i>	905.05	861.34	47.30	43.03	–	–

in *Full Info*.¹⁶ Differences between *Industry* and *No Info* and these three treatments are not only sizable (e.g., action *e*, corresponding to the Walrasian quantity, is chosen in *Selection* more than three times more than in *No Info* in the last 30 rounds), but significant in every relevant comparison.¹⁷

Table 3 complements Fig. 3 by focusing on Cournot and Walras quantities, characterizing participants' behavior when exposed to social information. The first two columns in Table 3 show the average payoffs obtained in the experiment when playing Cournot Nash (action *c*) and the Walrasian quantity (action *e*). Note that the Cournot-Nash quantity is more profitable than the Walrasian quantity in all five treatments. If decisions were driven by absolute payoffs, one would expect the switching ratios from the Cournot-Nash quantity (i.e., switching from Cournot to any other action) to be smaller than the switching ratios from Walras (i.e., switching from Walras to any other action) in the treatments, particularly when the difference is larger (as in *Market*, *Full Info* and *Selection*). As columns 3 and 4 in Table 3 document, this is only observed in *No info*, when participants only receive information about their own payoff. Once information about relative performance in the market is provided, a reversal of the switching ratios is observed.

How does information about average market profits push subjects to change switching probabilities? An intuitive answer is that this information leads them to make comparisons. In the last two columns in Table 3 we see how many times (as a percentage) Cournot-Nash and Walras actions yield profits below the social aspiration level (i.e., the average market profits in *Market*, *Full Info* and *Selection*). In *Market*, subjects playing Cournot-Nash Cournot earned below the average market almost half of the times (48.66%). In *Selection* and *Full Info* the percentage goes up to more than 60% and 65%, respectively. In other words, and in line Morales and Fernandez-de-Cordoba (2012), Walras never makes participants to fail reaching their aspiration profit level, but Cournot does quite frequently. Those participants taking the average market profit

¹⁶ These differences with the other two treatment are even more prominent in the second half of the experiment.

¹⁷ The Walrasian quantity is chosen significantly more frequently in *Market*, *Full Info* and *Selection* than in *No info* (*p* values < 0.000 in all cases, running logistics regressions). Table 9 in the appendix shows that the probability of choosing the Walrasian quantity increases 21 pp in *Market*, 29 pp in *Full Info*, and 25 pp in *Selection*, relative to *No Info*. Differences between *Market*, *Full Info* and *Selection* are not significant running Wald tests and using Bonferroni adjusted *p*-values.

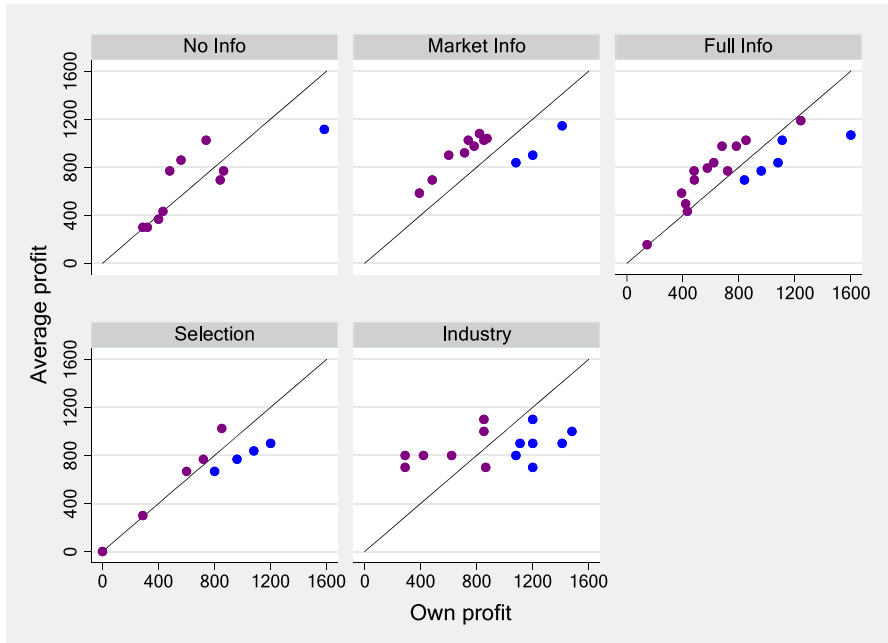


Fig. 4 Switching behavior as a function of one's own payoff and average payoffs by experimental treatment. Colors identify switching intensity, from low to high, using a blue-red scale

as their aspiration profit level have reasons to switch away from Cournot-Nash, but never from Walras.

Figure 4 displays switching behavior by experimental treatment as a function of individual own profit and the average profit. Each circle represents a pair of one's own payoffs in the horizontal axis and average market payoffs in the vertical axis (average industry payoffs for *Industry*).¹⁸ Colors identify switching intensity in a blue-red scale: dark blue for switching ratios below 0.25; light blue for switching ratios between 0.25 and 0.50; light red for switching ratios between 0.5 and 0.75 and finally dark red for switching ratios larger than 0.75. The main diagonal serves as a reference line: circles below the main diagonal correspond to combinations where one's own payoff is above average payoffs.

As participants were not informed about average payoffs in *No Info* one would expect switching behavior not to depend on whether the pair is located below or above the main diagonal and this is what we observe: red and blue circles appear both below and above the main diagonal in *No Info* (top left panel). However, the

¹⁸ We impose an arbitrary threshold limiting which circles are shown in Fig. 4: only those with at least 10 observations in the experiment (a 0.37% frequency). We admit the threshold $n = 10$ is arbitrary, but it is a way to exclude observations with extremely low frequencies that are not representative of the average behavior. The observations included in the graph represents more than 96% of total observations for all treatments, except for treatment *Industry*, which represents 66%. This last treatment is *peculiar* because most pairs correspond to unique observations. In Appendix C we include the graph with a threshold $n = 5$; for the treatment *Industry*, the graph represents 83.27% of all observations.

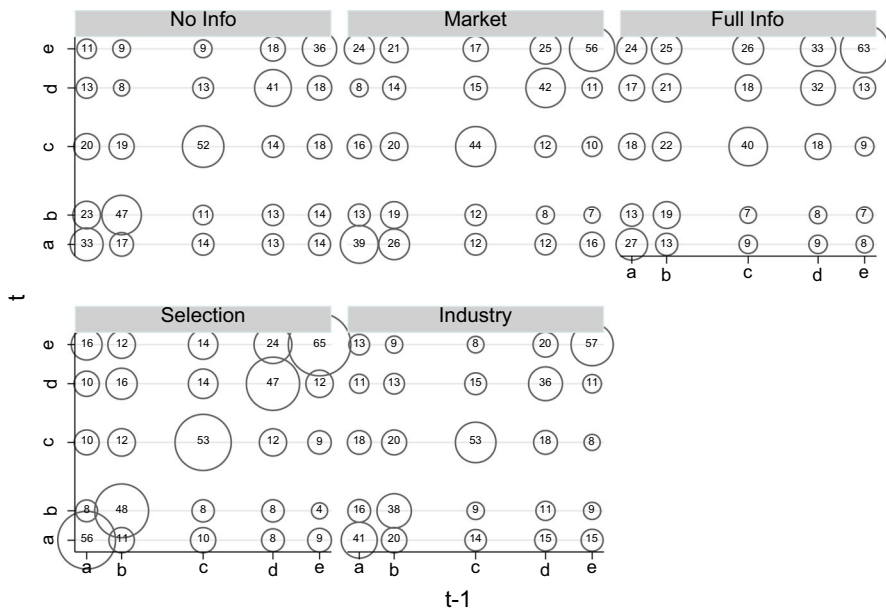


Fig. 5 Switching behavior by experimental treatment. Switching behavior from $t-1$ (horizontal axis) to t (vertical axis) for the five treatments. Circles sizes represents column's frequency (column numbers add 100)

main diagonal becomes relevant for the other treatments: red dots concentrate above the main diagonal and blue dots below the main diagonal, suggesting that participants switch more frequently if profits are below their aspiration level when making a choice for the next period.

Figure 5 presents data on which actions are chosen to switch in each treatment by participants. Table 8 in Appendix E shows the corresponding numerical transition matrixes. In the horizontal axis of Fig. 5 we display the quantity chosen in $t-1$ and in the vertical axis the quantity chosen in t , being the numbers in each column the relative frequency of actions chosen in t for each quantity chosen in $t-1$. Hence, for each column, all numbers add up to 100. To help the reader visualizing the switching patterns, we include circles with a size proportional to the frequency.

In *No Info* and *Industry*, switching towards higher quantities when producing little (playing actions a, b and c) and switching towards smaller quantities when producing a lot (playing actions e, d and c) is quite symmetric. In *No info* switching to produce more from a, b, and c happens 67, 36 and 22% of the times, while switching to produce less from actions e, d and c 64, 40 and 25%.¹⁹ However, when participants receive information about market profits, participants exhibit a preference for switching towards higher quantities. In *Market*, switching to produce more from actions a, b and c happens 61, 55 and 32% of the times, and switching to produce less from actions e, d and c 44, 32 and 24%. The asymmetric switching pattern is even more

¹⁹ In *Industry*, the frequencies are 59%, 42% and 23% and 43%, 44% and 23%, respectively.

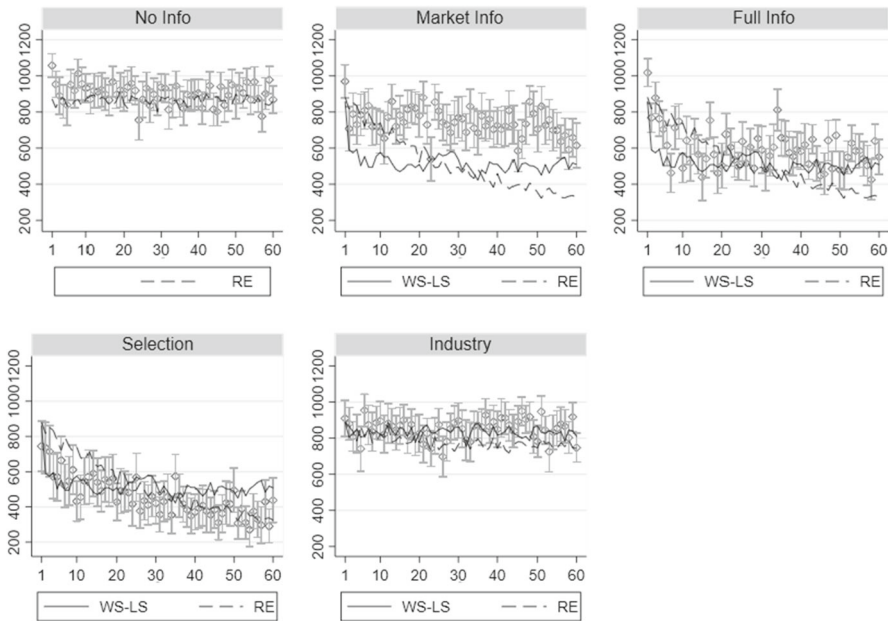


Fig. 6 Average profits of reinforcement learning rules (with confidence intervals, over time, for each treatment)

pronounced in *Full Info*: switching from a, b and c happens 73, 68 and 44% of the times and switching to lower quantities from e, d and c 34, 35 and 16%.²⁰

Switching behavior hits profits. Figure 6 displays average profits (and confidence intervals) over time in each treatment, together with simulated average profits for both reinforcement learning rules, as a benchmark. Profits remain flat around 900 (Cournot-Nash payoffs) when average market profit is not provided to subjects (as in *No Info* or *Industry*) and profits decline over time when average market profits are given to them (as in *Market*, *Full Info* and *Selection*). In *Selection*, the sharp profit decline is in line with our simulations for the RE rule. *Market* gets closer to our simulations for the WS-LS model, with an initial decline in profits to a stable plateau (and a final drop in the last ten periods).²¹ *Full Info* displays a pattern between *Selection* and *Market*, as it exhibits a larger initial decline to stabilize later in the experiment. We summarize these findings in the following result:

Result 4. *Switching behavior is significantly shaped by average market profits, as participants switch away from small quantities to large ones more often when their performance is below their (social) aspiration level, driving prices and payoffs down.*

²⁰ In *Selection*, 44%, 40% and 28% and 35%, 28% and 18%, respectively.

²¹ The level at which average profits stabilize depends on the noise parameter in the same way that average prices do.

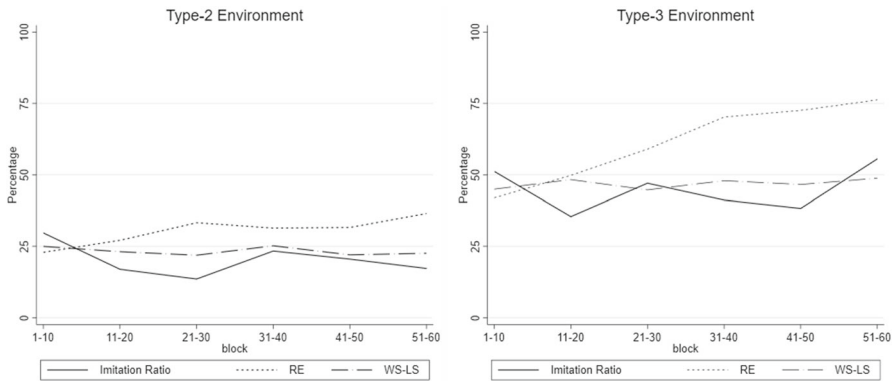


Fig. 7 Imitation ratios and reinforcement learning rules' predictions (Market treatment, by blocks of ten rounds)

5 Social learning and imitative heuristics

So far, our experimental results show how information about average profits push subjects towards higher quantities and markets towards lower prices. Still, subjects could have followed an imitative heuristic because they were always informed about the action profile in their market.²² For some combination of actions, or profiles, imitating the best is still straightforward. Consider for example an action profile with two different quantities, because two participants choose the same quantity. We will call these cases type-2 environments. By design, one action will yield payoffs below the average market profit, and it is immediate to find out that the other action is the most successful (the “best”) action in the market.²³ With three different actions (type-3 environments), the signal is still partially informative, as a participant earning below the average payoff can infer that one of the other two actions yields the best payoff in the market. Type 2 environments are quite frequent in our dataset, as in *Market* we find 466 type 2 profiles (51.78%) and 380 profiles of type 3 (42.22%). The remaining 6% are instances in which all three players played the same action.

If below the average payoff, imitation implies switching with probability 1 in type-2 environments and to one of the two observed actions in type-3 environments. Consequently, we define the imitation ratio as the fraction of times that if below the average market profit, subjects switch to the other observed action in type-2 environments and to one of the two observed actions in type-3 environments. Imitation ratios capture the extent to which subjects switch in a manner consistent with the imitate the best heuristic. Figure 7 displays the evolution of imitation ratios by blocks of ten rounds in these two types of environments (bold solid line), and the imitation ratios computed from our simulations for the WS-LS and RE reinforcement rules (with social aspiration).

²² We thank Robert Sugden for his insistence on this issue during a seminar.

²³ The literature considers other imitative rules in which information in absolute terms is needed, as for example the proportional imitation rule of Schlag (1998) and Morales (2002). But these rules cannot be applied in our context, as the individual information about the rivals' payoffs wasn't revealed to subjects.

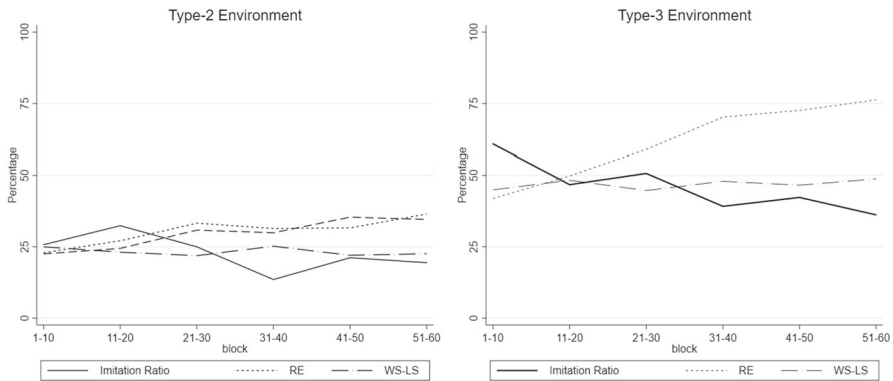


Fig. 8 Imitation ratios and reinforcement learning rules' predictions (Full Information treatment, by blocks of ten rounds)

For the WS-LS rule it is easy to compute the “theoretical” imitation ratios, as under this rule, subjects choose randomly when dissatisfied with their actions. This immediately implies that the imitation ratio in type-2 environment is 25% (when switching away, there is one chance out of 4 of choosing the observed action played by the “best” opponent) and is 50% in type-3 environment (the chances of choosing one of the two observed actions are 2 over 4). For the learning rules with more general state space, switching behavior depends of the mixed strategy. This is why we computed the imitation ratios from the simulations. For the type-2 environments, the computed imitation ratios are slightly above the 25% level of the WS-LS while spans from 50% to nearly 75% for type-3 environments.

The left panel in Fig. 7 presents data from type-2 environment. Imitation ratios oscillate around its average value, 21.33%, strikingly close to the reinforcement rules with social aspiration level predictions. In the right panel, with data from type-3 environments, we again observe that the imitation ratio oscillates around its average value, 47.24%, in fact in the lower bound of the imitation ratios derived from reinforcement rules with social aspiration levels. Our conclusion is that genuine imitation plays a limited role in explaining the experimental results, as imitation ratios are not different from what reinforcement rules with social aspiration level predict.

The limited role played by imitation in *Market* (when subjects are provided with social information) may be explained in at least two ways. First, social information makes subjects to disregard imitative heuristics, as social learning heuristics becomes more salient. Second, subjects are not able to correctly imitate the best because they have to compute which action yields the largest market profit and this is a cognitively demanding task.²⁴

Our *Full Info* treatment may thread light on both explanations because participants are given both the market payoff and the payoffs and actions profile of all participants in their market. Figure 8 below displays the imitation ratios in *Full Info*. In both type-2 and type-3 environments, imitation ratios are again around the lower bound of the

²⁴ We thank one referee for pointing this out.

predictions from reinforcement learning rules with social aspiration levels. Putting together Figs. 7 and 8, we may see that imitation ratios are slightly above in *Full Info* than in *Market*, especially during the first blocks of the experiment (during the sharp decline in average prices and profits discussed above). From Block III on, imitation ratios stabilize at levels like those observed in *Market*.

How should we interpret the low imitation ratios presented? Even when running a horse race between imitation heuristics and social learning heuristics to determine a winner is a futile exercise, and far from our goals in this paper, we cannot rule out a combined effect of both heuristics. Our results strongly support this conjecture: prices in *Full Info* are below those observed in *Market*, (and significantly so, see Table 2), the Walrasian quantity is chosen more frequently in *Full Info* (see Fig. 3), and profits are consequently lower (as suggested by Fig. 6).

We may still run an illustration of how this combination works by taking a closer look to situations in which these two types of heuristics –social comparison and imitation– makes different predictions. We focus on type-3 environments in which the firm choosing the second highest quantity gets the second highest profit but still above the market average payoff. This firm may adopt the recommendation from the social comparison rule (i.e., reinforce the current choice) or the one recommended by imitative heuristics (i.e., switch to the action yielding the highest payoff in the market).

Table 4 below presents data on how firms in this situation choose between imitating the best (as they can tell that another firm is doing better) and reinforcing their choice (as they are still above the market profit, and the social learning heuristics recommends not to switch). In Table 4 we can see the frequency of each action (following one heuristic or the other), as taken by the firm with the second highest profit in the market. Table 9 in the Appendix shows the frequencies of each action profile in which both models collide (five in total). In *Full Info*, we observe this scenario 192 times, slightly below 8% of all times. We also include the number of times subjects follow a simple myopic best response heuristic, choosing a best response to the actions observed in the previous period (as a benchmark).

Table 4 Behavior in Type-3 environments Frequency of imitation, social learning, and best response behavior

	Frequency	%
Social learning	58	30.2
Imitation	66	34.3
Best response	23	11.9
Any other	17 (50)	8.7*
Total	192	— **

Case in which social comparison and imitation yield different predictions in Type-3 environments

*Any other refers to how frequently each other action (other than the ones predicted by the three models above, two or three depending on the case) is chosen on average

**As the best response prediction coincides with the imitation one in some cases, the sum of percentages adds more than 100%

Table 4 strongly suggests that in cases in which participants may identify an action yielding the best payoffs in the market, social learning prediction (e.g., reinforcing the current choice, and not switching) is still followed as frequently as the action recommended by imitation heuristics. Both Imitation and social learning heuristics predict better than myopic best response (p value < 0.000 , using a exact Fisher test counting success and failures of each model), but no significant differences are observed between these two recommendations (p value $= 0.445$).

Result 5. Even in situations in which imitation is simple to learn, when imitation learning models and social learning models make different predictions, imitation is not followed more frequently than social learning by participants.

6 Conclusions

In the context of Cournot competition, a standard mechanism incorporating information about the choices and success of others is imitation of the most successful player. The market is predicted to move away from the Nash prediction in the direction of the Walrasian outcome under this rule, and this prediction has been extensively confirmed by lab experiments (Apesteguia et al. 2007, 2010).

In this paper, we study an alternative heuristic available to players when they are provided with aggregate information about market performance: *reinforcement with social aspiration level*. This heuristic positively reinforces current choices when players perform relatively better than the market, and they are negatively reinforced otherwise.

This paper offers several contributions. First, we prove some convergence results to the competitive outcome for three classes of reinforcement learning when the aspiration level is set to the average profit in the market: “Win-Stay, Lose-Switch” rule (Thorndike 1911), and the rule in Roth and Erev (1995), Erev and Roth (1998). These results contrasts with the findings by Oechssler (2002) that proved convergence to Collusion when the aspiration level is the average payoffs in the industry (where an industry is composed of several markets). Our results highlight how sensitive learning rules are to the specification of learning process.

Second, we investigate whether the asymptotic properties of the reinforcement rules are relevant for the short run description of the market dynamics. In a simulation exercise, we show that these reinforcement rules converge fast to the competitive outcome, as shown by simulations run for 60 periods. However, when the aspiration level is set to the average industry profit, the “Win-Stay, Lose-Switch” rule lingers around the Nash equilibrium price unless the error probability is almost zero. Thus, for this case, asymptotic properties are not a good description of the short run behavior of the market. Also, for the RE rule, behavior stays close to the Nash prediction.

We finally test the social comparison heuristics in a stylized experiment run under identical conditions to previous studies on imitation. The only remarkable change is that in some treatments participants do not receive information about individual profits and do get information about the average performance in the market. Two main conclusions can be drawn from our research.

First, although the *imitative* and the *social comparison* heuristics are quite different, they share the same *curse* in Cournot markets. Under both heuristics, players converge

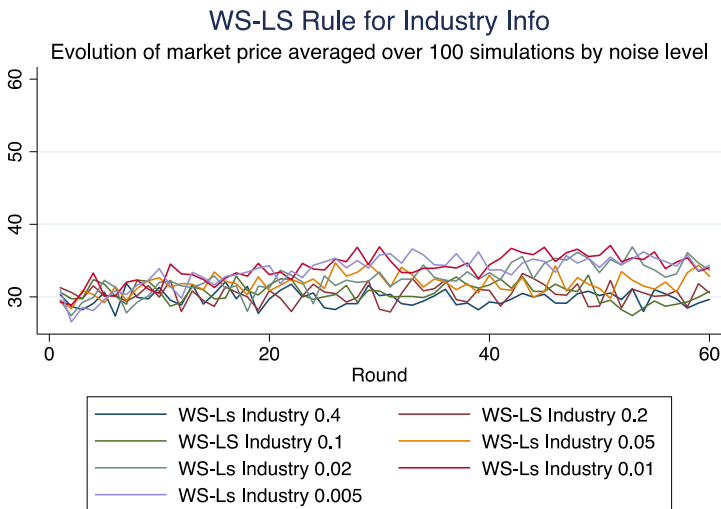
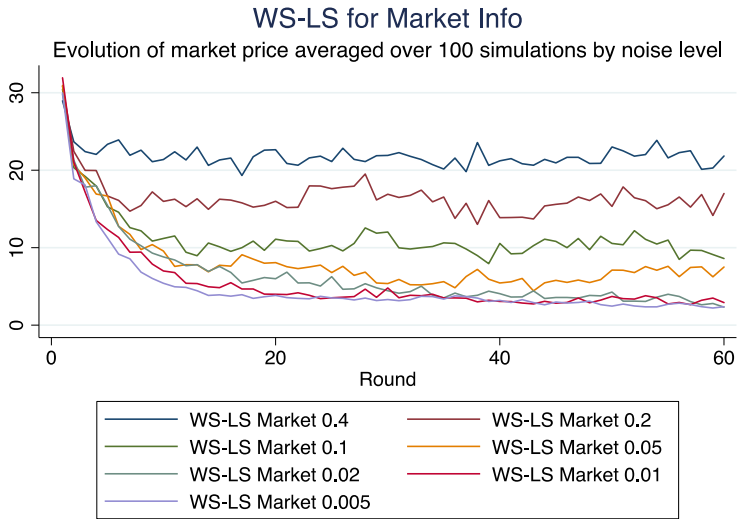
to the worst outcome for players: the Walrasian solution. Players in the short run (less than 60 periods) seem to be unable to efficiently handle the information conveyed in the choices of others. By adopting the *social comparison* heuristic, players do try to beat the market, and reinforce actions when their profits are above the average market profits. The Walrasian quantity has an interesting property: by choosing it, they never get profits below the average market profits. Hence, it is always positively reinforced. In this sense, our results confirm our hypotheses that the asymptotic properties of social reinforcement rules are relevant for the short run description of the dynamics of the markets. We also confirm the predictions based on simulation results when experimental subjects are informed about the average profits in the industry. Rather than converging to collusive practices, as the theoretical asymptotic result would prescribe, behavior stays around the Nash prediction.

Second, participants consider imitative and social comparison heuristics to be at least partially orthogonal. Some players use the *social comparison* rule and do no further exploit the available information to identify the best action in the market and imitate it. Our results strongly suggest that individuals combine these heuristics in our experiment, following a pattern consistent with social comparisons (when individual actions are not observable, as in *Market*) and consistent with both heuristics when the actions and profits profile are observable by subjects (as in *Full Info*). Interestingly, a small amount of information generates a major behavioral change. Even when market prices and profits are driven closer to the Walrasian outcome when both types of information are available, as in *Full Info*, by comparing themselves with other individuals beat the market, and by using the social comparison heuristics cannot *beat* Walras. Even in cases in which both models make different predictions, the recommendation of social learning (do nothing and reinforce your action) is as popular as the recommendation of imitate the best, when a better action is apparent.

Our experiment has also some limitations. Even when subjects are given the chance to learn how to play the game over 60 rounds, we cannot reach any sensible long-run conclusions. It would be interesting to apply the methodology develop by Friedman et al. (2015) to extend competition in the very long-run (e.g., 1000 periods) to see whether firms, after converging to Walras in 50–60 periods, escape from this trap and learn to play more profitable and collusive actions.

Acknowledgements Antonio acknowledges financial support from research project PY20-00069. Ainhoa acknowledges financial support from Spanish Ministry of Science and Innovation (PID2021- 123053OB-I00), Generalitat Valenciana (AICO/2021/005), and Jaume I University (UJI- B2021-23).

Appendix A. Sensitivity analysis to noise level parameter



Result A. For the Market Information, market average prices quickly drop over the first few rounds –competitive price is 0– and the noise parameter sets the level which the market price stabilizes around: the lower the noise parameter, the lower that price level. For the Industry Information, we only find a slightly upward trend towards collusive prices –collusive price is 60– when the noise parameter is below 1%.

Appendix B. Estimating learning rules

In Table 5 we present the outcome of probit models with random effects to analyze switching behavior

Models in Table A closely follow the reinforcement learning literature and consider as independent variables the current own payoff, *Own profit*, and the relative propensity of the chosen action, *Relative propensity*, defined as the sum of all past payoffs yielded by the action, divided by the sum of all past payoffs, and the dummy *Lagged exclusion*, to control for exclusion in the previous period in *Selection*. The coefficients for both variables are significantly negative in all models, in line with the theoretical intuition: the larger the individual payoff or its relative propensity, the larger the probability that a subject will play the same action in the next period.

Table B explicitly incorporates social comparison as an explanatory variable by including a new explanatory variable: *Relative performance*, as the distance between the aspiration level (average profit in the market or in the industry) and the individual profit in each period t . We observe that the estimate of the variable *Relative performance* is positive and highly significant in all treatments. The positive and significant effect of *Lagged exclusion* in model (10) is in line with the intuition that those excluded from the market in the previous round experiment more, and change their decisions more often.

Table C distinguishes between two interesting cases: when the decision maker is below (with the dummy *Below*) and above (with the dummy *Above*) the social aspiration level. It is not difficult to see that social comparison has a strong asymmetric effect. In all treatments, the dummy variable *Below* is positive and highly significant, while the dummy variable *Above* is only weakly significant for treatment *Selection*.

Result B. Reinforcement terms are negatively correlated with the switching probability in all treatments. Social comparison plays a significant role when relative performance info is provided (either at the market or the industry level), but the effect is stronger for those individuals below the market. Selection reinforces the effect of social comparison, and those excluded from the market switch away with higher probabilities.

Table 5 Probability models

(A)

	<i>No info</i>	<i>Market info</i>	<i>Full market info</i>	<i>Market selection</i>	<i>Market selection</i>	<i>Industry info</i>
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Own profit</i>	− 0.290 (0.04)***	− 0.229 (0.03)***	− 0.069 (0.03)**	− 0.286 (0.02)***	− 0.082 (0.03)***	− 0.408 (0.04)***
<i>Relative propensity</i>	− 706.018 (51.64)***	− 507.406 (52.74)***	− 557.154 (51.61)***	− 376.329 (48.93)***	− 396.569 (49.53)***	− 592.314 (50.9)***
<i>Lagged exclusion (dummy)</i>					288.044 (28.43)***	
<i>Log likelihood</i>	− 1451.149	− 1469.171	− 1608.350	− 1461.467	− 1416.676	− 1416.337
<i>Observations</i>	2655	2655	2655	2640	2640	2655

(B)

	<i>Market information</i>	<i>Full market info</i>	<i>Market selection</i>	<i>Market selection</i>	<i>Industry information</i>
	(7)	(8)	(9)	(10)	(11)
<i>Own profit</i>	− 0.080 (0.04)**	0.093 (0.03)***	− 0.093 (0.03)***	0.066 (0.04)*	− 0.174 (0.08)**
<i>Relative propensity</i>	− 516.426 (53.26)***	− 506.944 (52.34)***	− 389.572 (49.36)***	− 407.866 (49.79)***	− 591.667 (51.06)***
<i>Relative performance</i>	0.558 (0.07)***	0.916 (0.08)***	0.813 (0.11)***	0.702 (0.011)***	0.284 (0.09)***
<i>Lagged exclusion (dummy)</i>				265.008 (29.04)***	
<i>Log likelihood</i>	− 1440.077	− 1543.068	− 1430.021	− 1393.120	− 1411.2254
<i>Observations</i>	2655	2655	2640	2640	2655

(C)

	<i>Market information</i>	<i>Full market info</i>	<i>Market selection</i>	<i>Industry information</i>
	(12)	(13)	(14)	(15)
<i>Own profit</i>	− 0.1658 @ (0.03)***	− 0.037 @ (0.03)	− 0.123 @ (0.03)***	− 0.209 @ (0.05)***
<i>Relative propensity</i>	− 508.2352 @ (53.16)***	− 493.698 @ (50.4)***	− 407.8962 @ (49.78)***	− 598.63 @ (51.12)***

Table 5 (continued)

(C)

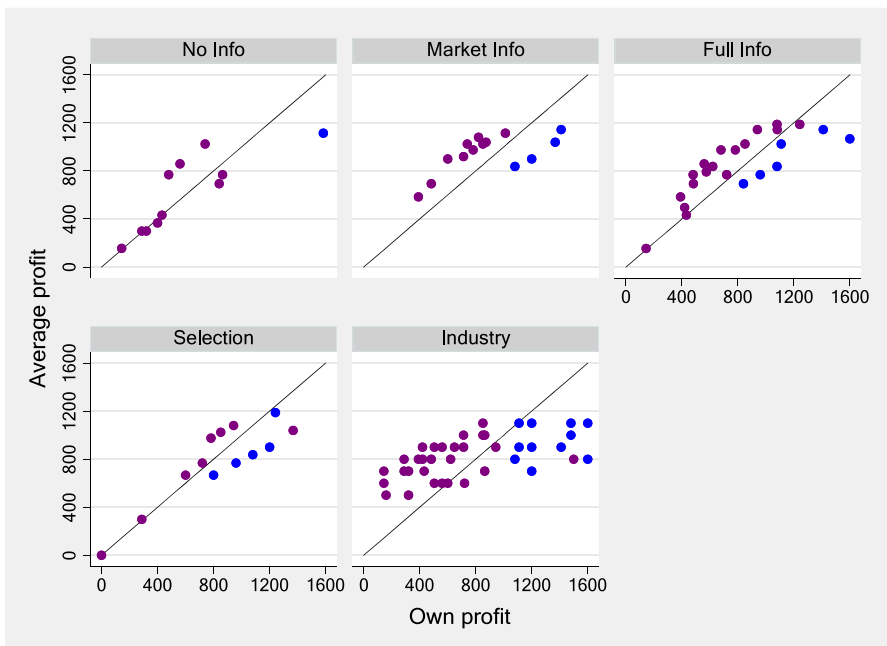
	Market information (12)	Full market info (13)	Market selection (14)	Industry information (15)
<i>Below (dummy)</i>	136.2407 @ (39.64)***	274.010 @ (33.12)***	248.9168 @ (33.48)***	182.167 @ (30.44)***
<i>Above (dummy)</i>	− 42.0775 @ (41.02)	12.149 @ (36.71)	− 68.2063 @ (36.46)*	
<i>Log likelihood</i>	− 1443.029	− 1535.016	− 1393.120	− 1398.836
<i>Observations</i>	2655	2655	2640	2655

All coefficients and standard errors are multiplied by 10^3 (Standard errors are in parenthesis)

Only cases with $s_i^{t+1} \neq s_i^t$ considered. Marginal effects after random-effects probit regressions

***1% level, ** 5% level and *10% level

Appendix C. Switching behavior with threshold value $n = 5$ by experimental treatment



Appendix D. Börgers and Sarin (2000) reinforcement rule with aspiration level in Cournot games

Simulations

When the aspiration level is set to the average payoff in the game, a full characterization of the convergence results of reinforcement learning rules is feasible for a 2×2 version of the Cournot game, where the two actions are the Nash equilibrium q^E and the Walrasian quantity q^W . For this example, as well as in the simulations and experimental sessions, we will use the parametrization in Apesteguia et al. (2007): costless production and linear demand $P = 120 - Q$. For a duopoly, we have $q^E = 40$ and $q^W = 60$. The payoff matrix of the reduced Cournot game is the following

	q^E	q^W
q^E	1600, 1600	800, 1200
q^W	1200, 800	0, 0

The Nash equilibrium action is strictly dominant. The following two observations are the key to the convergence results away from the Nash equilibrium action.

Lemma 1 *For every reinforcement rule that sets the average payoff as aspiration level, the Nash equilibrium output q^E will never be positively reinforced and sometimes it will be negatively reinforced.*

Lemma 2 *For every reinforcement rule that sets the average payoff as aspiration level, the Walrasian output q^W will never be negatively reinforced and sometimes it will be positively reinforced.*

Lemmas 1 and 2 immediately imply starting from an arbitrary initial state, along any path of the dynamic process, the system will take away probability from the collusive outcome and will increase the probability of the Walrasian outcome. As a result, the system will move in the direction of the Walrasian outcome.

In the convergence towards Walras, the system will display two features that generalize the results from the pure strategy learning rule analysed in the previous section. First, the probability of switching away from the chosen action will depend positively on how poorly the action performed compared to the market average. And second, for actions that are positively reinforced, there will always be a positive probability of the firm switching away from it in the next period (if the mixed strategy is not degenerated).

Consider the BS rule. Börgers and Sarin (2000) focus on decision problems with two actions, but we will work with the extension to m actions to produce simulations. The state of player i at time t is a mixed strategy $p^i(t) = (p_1^i(t), p_2^i(t), \dots, p_m^i(t))$

and actions are positively/negatively reinforced depending on whether the received payoff is above/below some endogenous aspiration level, that depends on the history of one's own payoffs.

We set the aspiration level to market average payoffs. Let $\hat{\pi}$ denote the largest profit that a firm can earn in the market game. Let α_i be the absolute value of the normalized reinforcement, i.e. $\alpha_i = \frac{1}{\pi} \left| \pi_i - \frac{\pi_1 + \pi_2 + \dots + \pi_n}{n} \right|$.

If firm i chooses action s at time t , and it pays off more than the market average payoff, then the mixed strategy is updated according to the following law

$$\begin{aligned} p_s^i(t+1) &= (1 - \alpha_i) p_s^i(t) + \alpha_i \\ p_{\tilde{s}}^i(t+1) &= (1 - \alpha_i) p_{\tilde{s}}^i(t) \text{ for } \tilde{s} \neq s \end{aligned}$$

If the payoff received is below the market average payoff, then the updated probabilities are

$$\begin{aligned} p_s^i(t+1) &= (1 - \alpha_i) p_s^i(t) \\ p_{\tilde{s}}^i(t+1) &= p_{\tilde{s}}^i(t) + \left(1 - p_{\tilde{s}}^i(t)\right) \frac{\alpha_i p_{\tilde{s}}^i(t)}{(m-2) + p_{\tilde{s}}^i(t)} \text{ for } \tilde{s} \neq s \end{aligned}$$

Unlike Börgers and Sarin (2000), we do not need to specify any updating rule for the aspiration level as the aspiration level is endogenously determined every period from the actions played by players.

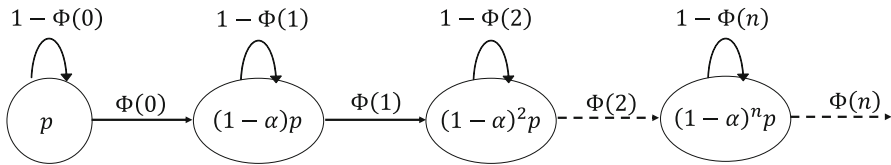
Consider now 2×2 Cournot game and any initial state. Lemmas 1 and 2 imply that the system will take away probability from the collusive outcome and will increase the probability of the Walrasian outcome. Next proposition shows that the system will go all the way towards the probability of choosing the Walrasian quantity converging to 1.

Proposition 2 *In 2×2 Cournot games, BS rule with social aspiration level converges with probability 1 to both players playing the Walrasian quantity.*

Proof Consider initial state with probabilities $p_E^1(0)$ and $p_E^2(0)$. The highly symmetrical nature of the learning rule implies that the ratio of $p_E^1(t)$ over $p_E^2(t)$ is constant for all time t and equal to $p_E^1(0)/p_E^2(0)$. The proof is by induction. Assume it is true for t and we will prove that it will be true for $t+1$. From t to $t+1$, the states remain the same if both firms have chosen the same strategy (and the claim would trivially follow). When they choose different quantities, states are updated. The firm that has played action E gets lower payoffs than the average and update $p_E^i(t+1) = (1 - \alpha_i) p_E^i(t)$ and $p_W^i(t+1) = (1 - \alpha_i) p_W^i(t) + \alpha_i$. The firm that has played action W gets higher payoff than the market average and update $p_W^j(t+1) = (1 - \alpha_j) p_W^j(t) + \alpha_j$ and $p_E^j(t+1) = (1 - \alpha_j) p_E^j(t)$. Because $\alpha_i = \alpha_j = \alpha$, it follows that

$$\frac{p_E^i(t+1)}{p_E^j(t+1)} = \frac{(1 - \alpha) p_E^i(t)}{(1 - \alpha) p_E^j(t)} = \frac{p_E^i(t)}{p_E^j(t)}$$

Therefore, starting from any initial state, for each mixed strategy of player 1, there is associated one and only one mixed strategy of player 2. It follows that the stochastic process can be described by a single Markov process with countable many states, where the label of each state is the probability that player 1 chooses the Nash equilibrium action E, and $\Phi(n)$ is the probability that the Markov process leaves the state $(1 - \alpha)^n p$.



Markov process associated to BS rule with social aspiration level

It is trivial to show that $\Phi(n) > 0$ for all n , because at each state, both players are characterized by a fully mixed strategy and randomization is independent across players (hence the probability that the actions of the two players are different is strictly positive). This probability is $(1 - \alpha)^n p [1 - (1 - \alpha)^n q] + [1 - (1 - \alpha)^n p] (1 - \alpha)^n q > 0$. All states are transient and that stochastic process will make player 1 (as well as player 2) to choose action E with a probability that converges to 0 as t goes to infinity. q.e.d

We finally show the simulation results for the BS rule in the figure above (Fig. 9). Together with the market and industry conditions, we also have simulations using the BS rule without aspiration levels to simulate situations in which the firms have

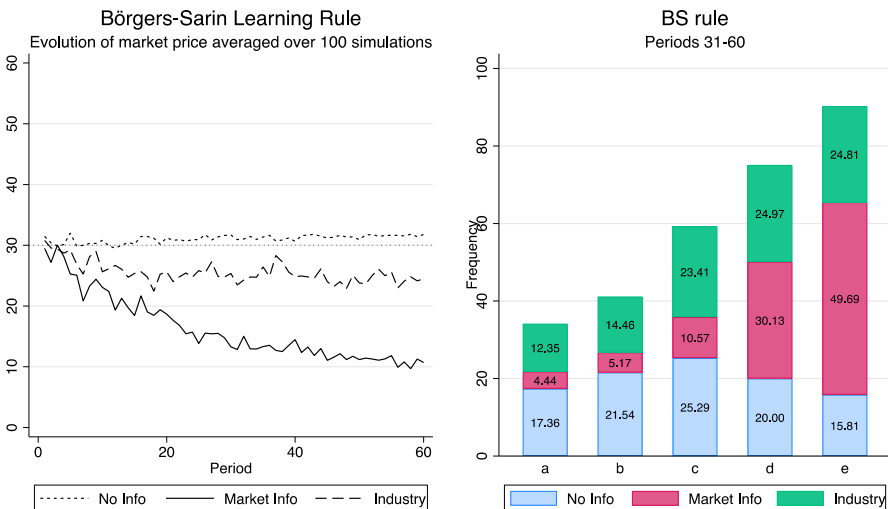


Fig. 9 Simulations from the Börgers and Sarin (2000) reinforcement rule. The left panel displays the simulation results from the Börgers and Sarin reinforcement rule, and the right panel displays the relative frequencies of quantities in rounds 31–60 for No Info, Market info and Industry treatment

no information about the rivals' performance. The parameters of the BS rule are the initial mixed strategy. The reported simulations correspond to the case in which the initial mixed strategy is selected randomly. Profits are divided by 1600 to normalize them.

Simulations for the three treatments start around the midpoint of the price interval at round 1. The BS rule without aspiration level yields market prices around the Nash equilibrium price over all periods. The distribution of actions for rounds [31, 60] is symmetric around the Cournot Nash action. When the BS sets the aspiration level to the market average payoff, market prices drop towards the competitive prices. The distribution of actions for rounds [31, 60] is skewed to the right, the Walrasian quantity accounts for half of the observations. Finally, the simulation for the Industry treatment show a slight decline of market prices. The distribution of actions for the second half of the simulation shows that actions c (Nash), d and e (Walras) represents 25% of the observations each.

Appendix E. Additional analysis

Treatment effects

	(1) Periods 1–60	(2) Periods 1–30	(3) Periods 31–60
Period	– 0.1077*** (0.02)	– 0.2323*** (0.05)	– 0.0752*** (0.03)
Industry	– 2.4611** (1.23)	– 3.4822*** (1.32)	– 1.4400 (1.56)
Market	– 6.9178*** (1.61)	– 6.5000*** (1.50)	– 7.3356*** (2.19)
Full info	– 13.2911*** (1.24)	– 12.9044*** (1.22)	– 13.6778*** (1.63)
Selection	– 12.4933*** (0.99)	– 9.2533*** (1.13)	– 15.7333*** (1.27)
Constant (No info)	35.8861*** (1.02)	36.9833*** (1.14)	35.2460*** (1.67)
Observations	1,500	750	750
Number of triopolies	25	25	25
	(1)	(2)	(3)
Wald Test p values*	Periods 1–60	Periods 1–30	Periods 31–60
Industry = No Info	0.1838	0.0333	1.0000
No Info = Market	0.0001	0.0001	0.0032
Market = Full Info	0.0006	0.0001	0.0163
Full Info = Selection	1.0000	0.0044	0.4661

Random-effects GLS regression with market prices (one observation per triopoly and period) as dependent variable. Robust standard errors in parentheses, clustered at the triopoly level

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

*Bonferroni-adjusted p values

Market prices per independent observation by experimental treatment

Period	Groups	Industry	No info	Market	Full info	Selection
First (t = 1)	1	38.00	47.00	30.33	31.33	39.33
	2	42.33	33.67	39.00	38.00	30.67
	3	14.00	44.67	33.67	41.67	36.33
	4	38.00	40.33	33.67	44.00	41.33
	5	36.00	42.67	53.67	41.67	36.00
	Average	33.67	41.67	38.07	39.33	36.73
All (1–60)	1	29.60	33.16	26.91	16.15	18.08
	2	33.29	34.63	27.29	18.13	21.22
	3	28.91	30.63	19.61	22.27	19.61
	4	27.45	30.40	27.72	20.03	21.63
	5	31.45	34.18	26.88	19.97	20.01
	Average	30.14	32.60	25.68	19.31	20.11
First half (1–30)	1	29.07	35.36	29.08	17.13	22.59
	2	31.77	33.01	27.09	20.77	25.82
	3	28.88	31.26	22.09	22.66	22.43
	4	26.88	31.48	26.47	20.22	26.16
	5	32.91	35.81	29.69	21.61	23.64
	Average	29.90	33.38	26.88	20.48	24.13
Second half (31–60)	1	30.13	30.97	24.74	15.17	13.58
	2	34.82	36.26	27.50	15.50	16.62
	3	28.94	30.01	17.13	21.88	16.78
	4	28.02	29.32	28.98	19.84	17.10
	5	29.99	32.56	24.08	18.33	16.37
	Average	30.38	31.82	24.49	18.14	16.09

Choices in t as a function of choices in t – 1 by experimental treatment frequencies

t	t – 1				
Industry	a	b	c	d	e
a	41.32	20.29	14.48	14.72	15.14
b	16.32	37.68	9.4	11.45	9.18
c	17.71	19.57	52.7	17.76	7.99
d	11.28	13.29	14.95	36.45	10.71
e	13.37	9.18	8.47	19.63	56.97
N	576	414	649	428	588
No info	a	b	c	d	e
a	32.72	16.72	14.07	13.36	14.39
b	22.84	47.47	11.2	13.36	14.39
c	20.37	18.92	52.46	14.03	17.93
d	12.76	7.6	12.84	40.98	17.68

No info	a	b	c	d	e
e	11.32	9.29	9.43	18.26	35.61
N	486	592	732	449	396
Market	a	b	c	d	e
a	38.87	25.52	12.45	11.87	15.84
b	12.96	19.31	12.06	8.45	6.82
c	15.51	20.34	43.97	12.33	10.4
d	8.39	14.14	14.59	42.01	10.87
e	24.27	20.69	16.93	25.34	56.07
N	548	290	514	438	865
Full info	a	b	c	d	e
a	26.76	13.11	8.95	8.96	7.78
b	13.38	19.26	7.16	7.94	6.8
c	18.39	21.72	40.16	17.72	9.21
d	17.39	21.31	18.09	32.18	12.79
e	24.08	24.59	25.65	33.2	63.42
N	299	244	503	491	1,118
Selection	a	b	c	d	e
a	55.8	10.59	10.41	8.44	9.2
b	7.79	48.47	8.12	8.15	4.16
c	10.17	12.47	52.83	12.23	8.93
d	10.49	16	14.4	47.45	12.29
e	15.74	12.47	14.24	23.73	65.43
N	629	425	653	687	1,131

Collision of social and imitative learning models

Action profiles	# of observations
Frequencies	
a c d	19
b c d	20
a d e	46
b d e	36
c d e	71
Total	192

Walrasian quantities

	(1)	(2)
	Periods 1–60	Periods 31–60
Marginal effects		
Period	0.0016*** (0.00)	0.0016* (0.00)
Industry	0.0893** (0.04)	0.0657 (0.05)
Market	0.2062*** (0.03)	0.2209*** (0.05)
Full Info	0.2868*** (0.04)	0.3129*** (0.05)
Selection	0.2533*** (0.02)	0.3186*** (0.03)
Observations	13,500	6,750

Standard errors in parentheses

Marginal effects of logistic regressions

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Market prices by treatment

Periods	Industry	No info		Market		Full info		Selection
1	33.67 (11.23)	41.67 (5.10)		38.07 (9.26)		39.33 (4.96)		36.73 (4.04)
1–60	30.14 (2.28)	32.60 (1.98)	*** [+++]	25.68 (3.41)	*** [++]	19.31 (2.29)		20.11 (1.41)
1–30	29.90 (2.42)	** 33.38 (2.13)	*** [+++]	26.88 (3.00)	*** [++]	20.48 (2.08)	*** [++]	24.13 (1.76)
31–60	30.38 (2.63)	31.82 (2.76)	*** [+++]	24.49 (4.57)	** [++]	18.14 (2.86)		16.09 (1.43)

Robust standard errors clustered at the triopoly level in parenthesis

Significance using Wald test and Bonferroni-adjusted p-values after random-effects GLS regressions. **at the 5% level, ***at the 1% level

Significance using Mann Whitney non-parametric tests (with five independent observations per treatment) ++at the 5% level, +++at the 1% level

References

- Alos-Ferrer, C., Ania, A.B.: The evolutionary stability of perfectly competitive behavior. *Econ. Theor.* **26**, 497–516 (2005). <https://doi.org/10.1007/s00199-004-0474-8>
- Alós-Ferrer, C., Shi, F.: Imitation with asymmetric memory. *Econ. Theor.* **49**, 193–215 (2012). <https://doi.org/10.1007/s00199-010-0554-x>
- Andreoni, J., Petrie, R.: Public goods experiments without confidentiality: a glimpse into fund-raising. *J. Public Econ.* **88**, 1605–1623 (2004)
- Apesteeguia, J., Huck, S., Oechssler, J.: Imitation: theory and experimental evidence. *J. Econ. Theory* **136**, 217–235 (2007)
- Apesteeguia, J., Huck, S., Oechssler, J., Weidenholzer, S.: Imitation and the evolution of Walrasian behavior: theoretically fragile but behaviorally robust. *J. Econ. Theory* **145**, 1603–1617 (2010)
- Azmat, G., Iriberry, N.: The importance of relative performance feedback information: evidence from a natural experiment using high school students. *J. Public Econ.* **94**(7), 435–452 (2010)
- Beggs, A.W.: On the convergence of reinforcement learning. *J. Econ. Theory* **122**(1), 1–36 (2005)
- Bendor, J., Mookherjee, D., Ray, D.: Aspiration-based reinforcement learning in repeated interaction games: an overview. *Int. Game Theory Rev.* **3**, 159–174 (2001a)
- Bendor, J., Mookherjee, D., Ray, D.: Reinforcement learning in repeated interaction games. *Adv. Theor. Econ.* **1**(1), 20011004 (2001b)
- Bicchieri, C., Fatas, E., Aldama, A., Casas, A., Deshpande, I., Lauro, M., Parilli, C., Spohn, M., Pereira, P., Wen, R.: In science we (should) trust: expectations and compliance across nine countries during the COVID-19 pandemic. *PLoS ONE* **16**(6), e0252892 (2021)
- Bilancini, E., Boncinelli, L.: Social coordination with locally observable types. *Econ. Theor.* **65**(4), 975–1009 (2018). <https://doi.org/10.1007/s00199-017-1047-y>
- Bilancini, E., Boncinelli, L.: The evolution of conventions under condition-dependent mistakes. *Econ. Theor.* **69**(2), 497–521 (2020). <https://doi.org/10.1007/s00199-019-01174-y>
- Blanes I Vidal, J., Nossol, M.: Tournaments without prizes: evidence from personnel records. *Manag. Sci.* **57**, 1721–1736 (2011)
- Blanton, H., Buunk, B.P., Gibbons, F.X., Kuyper, H.: When better-than-others compare upward: choice of comparison and comparative evaluation as independent predictors of academic performance. *J. Personal. Soc. Psychol.* **76**(3), 420–430 (1999)
- Börgers, T., Sarin, R.: Learning through reinforcement and replicator dynamics. *J. Econ. Theory* **77**(1), 1–14 (1997)
- Börgers, T., Sarin, R.: Naive reinforcement learning with endogenous aspirations. *Int. Econ. Rev.* **41**(4), 921–950 (2000)
- Bush, R.R., Mosteller, F.: *Stochastic Models for Learning*. John Wiley & Sons Inc. (1955)
- Buunk, B.P., Gibbons, F.X.: Social comparison: the end of a theory and the emergence of a field. *Organ. Behav. Hum. Decis. Process.* **102**, 3–21 (2007)
- Charness, G., Masclet, D., Villeval, M.C.: The dark side of competition for status. *Manag. Sci.* **60**(1), 38–55 (2013)
- Croson, R., Fatas, E., Neugebauer, T.: Reciprocity, matching and conditional cooperation in two public goods games. *Econ. Lett.* **87**, 95–101 (2005)
- Croson, R., Fatas, E., Neugebauer, T., Morales, A.J.: Excludability: a laboratory study on forced ranking in team production. *J. Econ. Behav. Organ.* **114**, 13–26 (2015)
- Cross, J.G.: A stochastic learning model of economic behavior. *Q. J. Econ.* **87**(2), 239–266 (1973)
- Deci, E.: *Intrinsic Motivation*. Plenum Press, New York (1975)
- Dijk, O., Holmen, M., Kirchlet, M.: Rank matters—The impact of social competition on portfolio choice. *Eur. Econ. Rev.* **66**, 97–100 (2014)
- Dixon, H.D.: Keeping up with the Joneses: competition and the evolution of collusion. *J. Econ. Behav. Organ.* **43**, 223–238 (2000)
- Duersch, P., Kolb, A., Oechssler, J., Schipper, B.C.: Rage against the machines: how subjects play against learning algorithms. *Econ. Theor.* **43**(3), 407–430 (2010). <https://doi.org/10.1007/s00199-009-0446-0>
- Erev, I., Haruvy, E.: Learning and the economics of small decisions. *Handb. Exp. Econ.* **2**, 638–700 (2013)
- Erev, I., Roth, A.E.: Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. *Am. Econ. Rev.* **88**, 848–881 (1998)
- Erikson, T., Poulsen, A., Villeval, M.C.: Feedback and incentives: experimental evidence. *Labour Econ.* **16**, 679–688 (2009)

- Etzioni, A.: Modern Organizations. Prentice-Hall, Englewood Cliffs (1971)
- Fallucchi, F., Fatas, E., Kölle, F., Weisel, O.: Not all group members are created equal: heterogeneous abilities in inter-group contests. *Exp. Econ.* **24**, 669–697 (2021)
- Fatas, E., Heap, S.P.H., Arjona, D.R.: Preference conformism: an experiment. *Eur. Econ. Rev.* **105**, 71–82 (2018)
- Fatas, E., Meléndez-Jiménez, M.A., Solaz, H.: Social hierarchies: a laboratory study on punishment patterns across Networks. *Econ. Inq.* **58**(1), 104–119 (2020)
- Fehr, E., Schmidt, K.: A theory of fairness, competition, and cooperation. *Q. J. Econ.* **114**, 817–868 (1999)
- Fischbacher, U.: z-Tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ.* **10**(2), 171–178 (2007). <https://doi.org/10.1007/s10683-006-9159-4>
- Fischbacher, U., Gächter, S.: Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *Am. Econ. Rev.* **100**, 541–556 (2010)
- Freidlin, M.I., Wentzell, A.D.: Random Perturbations of Dynamical Systems. Springer-Verlag, Berlin (1984)
- Friedman, D., Huck, S., Oprea, R., Weidenholzer, S.: From imitation to collusion: long-run learning in a low-information environment. *J. Econ. Theory* **155**, 185–205 (2015)
- Gortner, P.J., van der Weele, J.J.: Peer effects and risk sharing in experimental asset markets. *Eur. Econ. Rev.* **116**, 129–147 (2019)
- Hart, S., Mas-Colell, A.: Stochastic uncoupled dynamics and Nash equilibrium. *Games Econ. Behav.* **57**(2), 286–303 (2006). <https://doi.org/10.1016/j.geb.2005.09.007>
- Haruvy, E.: On the importance of relative payoffs in two-sided matching. *J. Inst. Theor. Econ.* **175**(1), 58–85 (2019)
- Hedlund, J.: Imitation in Cournot oligopolies with multiple markets. *Econ. Theor.* **60**, 567–587 (2015). <https://doi.org/10.1007/s00199-015-0878-7>
- Hedlund, J., Oyarzun, C.: Imitation in heterogeneous populations. *Econ. Theor.* **65**(4), 937–973 (2018). <https://doi.org/10.1007/s00199-017-1044-1>
- Izquierdo, L.R., Izquierdo, S.S., Gotts, N.M., Polhill, J.G.: Transient and asymptotic dynamics of reinforcement learning in games. *Games Econ. Behav.* **61**(2), 259–276 (2007)
- Kahneman, D., Tversky, A.: Prospect theory: an analysis of decision under risk. *Econometrica* **47**, 263–292 (1979)
- Kalai, E., Lehrer, E.: Rational learning leads to Nash equilibrium. *Econometrica* **61**, 1019–1045 (1993)
- Karandikar, R., Mookherjee, D., Ray, D., Vega-Redondo, F.: Evolving aspirations and cooperation. *J. Econ. Theory* **80**(2), 292–331 (1998)
- Lindner, F., Kirchler, M., Rosenkranz, S., Weitzel, U.: Social motives and risk-taking in investment decisions. *J. Econ. Dyn. Control* **127**, 104116 (2021)
- Morales, A.J., Fernandez-de-Cordoba, G.: The Walrasian output beats the market. *Int. J. Game Theory* **41**, 209–212 (2012)
- Morales, A.J.: Absolutely expedient imitative behavior. *Int. J. Game Theory* **31**, 475–492 (2002)
- Oechssler, J.: Cooperation as the result of learning with aspiration levels. *J. Econ. Behav. Organ.* **49**, 405–409 (2002)
- Palomino, F., Vega-Redondo, F.: Convergence of aspirations and (partial) cooperation in the prisoner's dilemma. *Int. J. Game Theory* **28**(4), 465–488 (1999)
- Pemantle, R.: A survey of random processes with reinforcement. *Probab. Surv.* **4**, 1–79 (2007)
- Posch, M., Pichler, A., Sigmund, K.: The efficiency of adapting aspiration levels. *Proc. r. Soc. Lond. Ser. B Biol. Sci.* **266**(1427), 1427–1435 (1999)
- Restrepo-Plaza, L., Fatas, E.: When ingroup favoritism is not the social norm a lab-in-the-field experiment with victims and non-victims of conflict in Colombia. *J. Econ. Behav. Organ.* **194**, 363–383 (2022)
- Roth, A.E., Erev, I.: Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term. *Games Econ. Behav.* **8**(1), 164–212 (1995)
- Rustichini, A.: Optimal properties of stimulus-response learning models. *Games Econ. Behav.* **29**, 244–273 (1999)
- Schlag, K.: Why imitate, and if so, how?: a boundedly rational approach to multi-armed bandits. *J. Econ. Theory* **78**, 130–156 (1998)
- Schoenberg, E.J., Haruvy, E.: Relative performance information in asset markets: an experimental approach. *J. Econ. Psychol.* **33**, 1143–1155 (2012)
- Thorndike, E.L.: Animal Intelligence. MacMillan, New York (1911)
- Tran, A., Zuckhauser, R.: Rank as an inherent incentive: evidence from a field experiment. *J. Public Econ.* **96**, 645–650 (2012)

Vega-Redondo, F.: The evolution of Walrasian behavior. *Econometrica* **65**, 375–384 (1997)

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.