J Econ Interac Coord (2006) 1:147–169
DOI 10.1007/s11403-006-0009-4

REGULAR ARTICLE

Can boundedly rational sellers learn to play Nash?

Roger Waldeck · Eric Darmon

Published online: 21 October 2006

Abstract How does Nash pricing compare to pricing with adaptive sellers using reinforcement learning (RL)? We consider a market game similar to Varian's model (Am Econ Rev 70:651–659, 1980) with two types of consumers differing by the size of their fixed sample search rule and derive the Nash search equilibrium (NSE) strategy (the density, the mean and the variance of the posted price distribution). Our findings are twofold. First, we find that the RL price distribution does not converge in a statistical sense to the NSE one except when competition is *à la* Bertrand. Second, we show that the qualitative properties of the NSE with respect to a change in buyers' search behavior are still valid for the RL distribution. The average price and the variance of both price distributions exhibit similar variations to a change in buyers' search.

Keywords Imperfect information · Price competition · Price dispersion · Search market equilibrium · Reinforcement learning · Numerical computation

JEL Classifications D43 · D83 · C63

R. Waldeck (✉)
GET ENST-Bretagne, Department LUSSI and ICI -UBO (EA-2652),
Technopole Brest Iroise CS 83818,
29238 Brest Cedex 3, France
e-mail: roger.waldeck@enst-bretagne.fr

E. Darmon
Faculté des Sciences Economiques et CREM,
CREM-CNRS (UMR 6211, University of Rennes 1 and CNRS),
7, place Hoche, 35065 Rennes Cedex, France
e-mail: eric.darmon@univ-rennes1.fr

1 Introduction

A large literature has been devoted to analyze market pricing in the presence of imperfectly informed buyers (cf. Stigler (1961) for a seminal paper). One usual setting is as follows: there are two types of buyers (informed and uninformed buyers) and the distribution according to these two types is ruled by an exogenous parameter. It is then generally assumed that sellers know ex ante the search characteristics of buyers. From this information, they can deduce the optimal profit-maximizing price distribution and in most cases, a dispersed price distribution arises as a consequence of the presence of two types of buyers. Varian (1980) is typical of this setting: some consumers are uninformed (and captive to one seller) while others are perfectly informed and have a fixed sample size search rule (i.e., visit the whole set of sellers). In that case, Varian (1980) shows that the unique symmetric equilibrium is in mixed strategy and he depicts the characteristics of the Nash Search Equilibrium (further NSE). Indeed, sellers' profit-maximizing strategy can be broadly comprehended as a mix between a high sales – low price strategy (thus targeting the informed population of buyers) on the one hand and a low sales – high price strategy (thus targeting the uninformed population of buyers) on the other hand. Waldeck (2006) points out that the link between information and pricing is often misleading in the context of search theory models. He shows that better information may lead to higher price dispersion and that more intensive search by informed shoppers may lead to higher average posted prices.

Although Varian (1980) suggests a dynamic interpretation based on the notion of sales, the NSE remains a static concept. Notably, these models fail to give a plausible account of how the choices made by less rational sellers compare to those made at Nash-equilibrium. In other words, they lack a theory of learning in games. Once we assume that sellers do not know buyers' search characteristics or that they have limited computational capabilities, sellers can no longer formulate a profit-maximizing price strategy and we suppose that they will rely on a simple reinforcement rule (adaptive learning) to establish a suitable price strategy. However, this choice can be disputed. At first view, the use of RL mechanisms has been justified by their ability to fit experimental data. In a seminal paper, Erev and Roth (1998) explore the robustness of the Nash predictions in normal form games with a unique mixed strategy equilibrium. For a large number of experiments in games that involve repeated plays of 100 periods, and those that are characterized by different information structures, they show that a simple one-parameter reinforcement model outperforms the equilibrium prediction of the game. Therefore, RL seems to find some support in the literature of experimental economics for explaining actual data.[1] Another argument in favor of reinforcement learning comes from Simon (1976). He argues

[1] Yet, as evidenced by Salmon (2001), one should remind that the econometric models used to retrieve agents' behavior may be inadequate and lead to false conclusions about the true underlying learning model. However, this restriction is inherent to many learning models and does not only apply to RL.

that agents confronted to a complex situation, may revert to simple heuristics to avoid computational overload. In addition, as noted by Simon, it is important to model *how* humans choose, i.e., procedural rationality and not just focus on defining the optimal nature of the outcome of choice, i.e., substantive rationality.

In this context, it is interesting to give a brief survey of the link between Nash predictions and outcomes resulting from the application of RL processes. We can identify two strands in the literature. The first one uses agent-based (ACE) models and RL processes to tackle issues related to market dynamics and organization. Kirman and Vriend (2001) propose an ACE model which tries to reproduce two stylized facts observed on Marseille fish market: high loyalty and high price dispersion. In their model, sellers are capacity constrained, prices are not posted, and buyers make repeated purchases. They show how buyers' behaviors become heterogeneous since sellers learn to discriminate buyers according to their loyalty. Using the same kind of behavioral model but in a capacity-unconstrained framework, Brenner (2002) shows how buyers and sellers may learn to bargain. As bargaining is costly, he determines in which cases, such a market converges to a posted price market.

A second approach asks whether the learning scheme converges to the Nash equilibrium predicted by standard game theoretical models. For instance, Posch (1997) and Beggs (2005) discuss the convergence to a mixed strategy equilibrium in 2×2 normal form games. Posch (1997) shows that reinforcement learning leads with positive probability to a cycling of strategy profiles. Beggs (2005) shows that the probabilities with which each agent plays each strategy converge to their equilibrium values almost surely. They both use a learning algorithm similar to the Erev and Roth's one, except that in Posch's model, reinforcements are re-scaled every period.[2] Hopkins and Seymour (2002) and Benaïm et al. (2005) show that NSE is dynamically unstable under a wide class of learning dynamics, including the reinforcement learning rule used by Erev and Roth or best response dynamic.

Our paper combines the two previous approaches. We are first interested in the issue of convergence towards Nash equilibrium. Yet, we consider this in the specific context of a search market and we use for that the same methodology (agent-based models and numerical simulations) as that developed by the first approach. However, the model developed here focuses on sellers' learning and then takes buyers' behavior as exogenously given.[3] The key finding of this paper is that even if convergence to the mixed strategy equilibrium is not observed, the Nash predictions with respect to a change in the market structure are still valid. Put differently, the reactions of adaptive sellers with respect to a change in the degree of information are qualitatively identical to those of Nash sellers.

---

[2] In their models, current fitness is equal to the cumulative past payoffs and the probabilistic choice rule is a simple proportional rule of current fitness (see also footnote 8 for the differences with our framework).

[3] Moreover, it is a simplistic view taken on the buyers' search behavior (only two types) mainly because analytical results for Nash exist in this case, thus enabling enables a comparison with RL.

Section 2 presents the simulation model. Section 3 introduces the NSE. Grounding on these propositions, we elaborate several hypotheses and present the protocol used to test them. Section 4 derives the main insights from the comparison between NSE and RL outcomes. Section 5 discusses the results and concludes.

2 The model

We model the market as a posted-price mechanism where S sellers (indexed by s) and B buyers (indexed by b) respectively produce and consume an indivisible and homogeneous item. Each market period $t (t = 1, \ldots, T)$ is divided into three sub-periods: first, sellers post prices; second, buyers visit sellers and transact; third, sellers evaluate the profit generated by the pricing rule currently used and reward this rule. Using the updated reward, they post a new price for period $(t + 1)$, etc.

2.1 Buyers

We deal with non-repeated purchases (no memory effect for buyers). There are $B = 1{,}000$ buyers. Each of them needs to purchase one unit of an indivisible good at maximum price v. This reservation price is common to all buyers and will further be set without loss of generality to 100. There are two types of buyers: *informed* buyers (in proportion a of the total population with $a \in [0, 1]$) and *uninformed* buyers. Informed buyers have fixed sample search strategy: they visit randomly $k > 1$ sellers at each period and buy at the lowest price; *uninformed* buyers just visit one seller and buy at that price.

2.2 Sellers

There are $S = 20$ sellers. There is no capacity constraint and sellers incur a constant *per* unit cost (c). As costs are identical for every seller, we further assume without loss of generality that $c = 0$. At the beginning of each period, sellers independently set the price they will post for this period. Posted prices are take-it-or-leave-it (no bargaining). Prices are discrete variables ranging from c to v with an exogenous step ϕ, sellers then have $(1 + (v - c)/\phi)$ price rules namely {Price c, Price $c + \phi$, Price $c + 2\phi, \ldots$, Price v}. Let us set $\phi = 1$, so that possible prices are simply $\{0, 1, 2, \ldots, 100\}$.[4] At each period, sellers decide which price (p_t) they will post by using a simple RL process.[5]

[4] We assume that sellers know the reservation price v rather than learn it. If not, prices higher than v would be rapidly eliminated through learning since these would generate a zero profit whenever posted.

[5] See Kirman and Vriend (2001) for more details about the learning process; see also Sutton and Barto (1998) for a general description of RL processes.

2.3 Reinforcement learning for price setting

Each price corresponds to a rule. In turn, each rule is defined by a triplet {Condition, Action, Fitness}. Rules have no condition part i.e., any price rule *can* be applied whatever the sellers' environment. The action part is simply the price sellers decide to post during the period. The fitness of a rule (F) measures its ability to satisfy sellers' objective (i.e. profit). At the first period, we assume that all sellers have the same expectation about the profit generated by all pricing rules. Hence, this initial expectation is price independent and all pricing rules are initially endowed with the same fitness noted F_0. This initial fitness F_0 can also be interpreted as sellers' expected belief about market-profitability.

$$F^i_{t=0} = F_0 = \delta(Bv)\ (\forall i) \quad \text{with } \delta \in [0, 1]. \tag{1}$$

With an *initial fitness coefficient* $\delta = 0$ (resp. $\delta = 1$), sellers are "pessimistic" (resp. optimistic) and expect a null profit (resp. maximal profit). During the subsequent periods, sellers revise the expected profit of each rule according to the actual instantaneous profits π_t generated by that rule. Let n_t denote the number of actual transactions implemented at period t. For any pricing rule i (used during period t), the fitness of rule i (F^i_t) is thus updated as follows:

$$F^i_{t+1} = F^i_t + \alpha\left(\pi_t - F^i_t\right) \quad \text{with } \pi_t := (p_t - c)(n_t). \tag{2}$$

Coefficient α $\left(\alpha \in]0, 1]\right)$ is *the reward updating parameter* and measures the relative weight attached to the most recent experience. This updating rule mimics an adaptive trial-and-error process: if, during the current period, the pricing rule i generates a higher profit than what it did in the past, its reward increases and hence, the probability of selecting this rule is reinforced for the next periods (see Eq. 3).[6]

Seller sets a price according to a trembling-hand process. In other words, sellers' behavior is characterized by a trade off between exploration and exploitation: at each period, each rule is selected with a probability drawn from a Boltzmann distribution (also called Logit distribution) described by Eq. 3:

$$\text{prob}\{\text{Select Rule } i\} = \frac{e^{\frac{\tilde{F}^j_t}{\tau}}}{\sum_j e^{\frac{\tilde{F}^j_t}{\tau}}} \quad \text{with } \tau > 0. \tag{3}$$

[6] Note that at period $t+1$, past profit π_{t-j} at period t–j accounts for $\alpha(1-\alpha)^j\pi_{t-j}$ in the current fitness.

$\widetilde{F}_t^i$ is the normalized fitness.[7] Parameter τ, called temperature, sets the trade off between exploration and exploitation: as $\tau \to 0$, sellers tend to select only the rules that have generated the highest payoffs in the past (choice of the "greedy action"). As τ increases, sellers tend to explore alternative rules more frequently tending to random choices as $\tau \to \infty$.[8]

3 Nash predictions and hypothesis testing

3.1 Theoretical predictions and hypotheses

The NSE leads to several propositions summed by Propositions 0 to 3. All the demonstrations are detailed in Waldeck (2006). The first proposition describes the equilibrium price distribution. The remaining propositions depict the variations of three main statistics on this market (average posted price, average price accepted by informed buyers and standard deviation of posted prices) with respect to buyers' search behaviors (i.e. parameters a and k). Concerning the Nash equilibrium, when all buyers are uninformed ($a = 0$), all sellers will charge the highest possible price, which is equal to the buyers' reservation price v. Indeed, undercutting the monopoly price will bring no additional sale and profits at any price below v may be increased by an increase in price since price elasticity is equal to zero. If all buyers are informed, competition by price undercutting will lead prices to $p = 1$ in the case of a discrete price grid.[9] A firm increasing its price from $p = 1$ will make no sale whereas a firm decreasing its price to $p = 0$ will make zero profit. In intermediate cases (mix of uninformed and informed buyers), each firm faces two markets: the captive market of uninformed consumers and the common market of shoppers. Without price discrimination, a firm randomizes its strategy between selling at a low price to both markets with a high probability and selling only to its captive market at a higher price.

Results related to NSE are summarized in Propositions 0 to 3 . For each proposition, we formulate a related hypothesis to test whether the RL-outcomes fit the NSE propositions.

---

7 We use a linear normalization of all fitness between 0 and 1: $\widetilde{F}_t^i = \frac{F_t^i - F_t^{\text{Min}}}{F_t^{\text{Max}} - F_t^{\text{Min}}}$ where $F_{t,s}^{\text{Max}}$ (resp. $F_{t,s}^{\text{Min}}$) is the maximum (resp. minimum fitness) observed by the seller at period t. Such transformation is neutral on the RL process since it does not change the relative ranking of the pricing rules. We made this transformation in order to avoid a calibration of the temperature parameter dependent on the absolute value taken by rewards and fitness.

8 They are two differences between our learning rule and the one used by Erev and Roth (1998). First, the updating of the fitness in equation [2] is done by averaging the past fitness with the current profit, whereas profits cumulate fitness in Erev and Roth's paper. Second, their probabilistic choice rule is a simple proportional rule of current fitness. We choose to stick to the logit model because of its axiomatic foundations issued from the psychological literature [e.g. De Palma and Thisse (1987) for a review].

9 In the case of a continuous support, the symmetric Bertrand equilibrium would be $p = 0$.

Proposition 0 (Price distribution) *The* (*symmetric*) *Nash pricing equilibrium distribution in such setting is as follows:*

- *Case with no informed buyers* ($a = 0$ *or* $k = 1$): *Monopoly pricing i.e.*, $p = v$ ($= 100$).
- *Case with no uninformed buyers* ($a = 1, k \neq 1$): *Bertrand competition i.e.*, $p = 1$.
- *General case with a mix of uninformed and informed buyers* ($0 < a < 1$, $k \neq 1$): $F(p; a, k) = 1 - \left(\frac{(1-a)(v-p)}{kap}\right)^{\frac{1}{k-1}}$ *with the lower* (*resp. upper*) *bound of the distribution equal to* $b(a, k) = (1 - a)v/((k - 1)a + 1)$ (*resp.* v).

Hypothesis 0 The RL price distribution converges to the Nash distribution.

Proposition 1 (Expected posted price) (i) *The NSE expected posted price decreases from the monopoly price v to the marginal cost 0 as a varies from* 0 *to* 1*;* (ii) *The NSE expected posted price increases with k* (*i.e.*, $k_1 < k_2 \Rightarrow E(p_1; a, k_1) < E(p_2; a, k_2)$).

Hypothesis 1.1 The RL mean price is non-increasing in a.

Hypothesis 1.2 The RL mean price is non-decreasing in k.

Proposition 2 (expected price paid by informed buyers) *The NSE expected price paid by informed buyers decreases* (i) *with a and* (ii) *with k.*

Hypothesis 2.1 The RL mean price paid by informed consumers is non-increasing in a.

Hypothesis 2.2 The RL mean price paid by informed consumers is non-increasing in k.

Proposition 3 (Standard deviation of posted prices) (i) *The standard deviation of the NSE posted prices is an inverse U-shaped function of a* ($0 < a < 1$). *The maximum of this function is approximately at*: $a = 70\%$ *for* $k = 2, a = 80\%$ *for* $k = 3, a = 90\%$ *for* $k = 4$. *This proportion increases for higher values of k and at* $k \geq 8$, *the maximum of the function is reached for values of a greater than or equal to* 99%. *In most cases*, *the variance is increasing with a* (*see Fig.* 1).

(ii) *Keeping a constant*, *the standard deviation of the NSE with respect to k is an inverse U-shaped function of k* (*Fig.* 2). *Whenever* $k < 8$, *the variance is increasing in k. The decrease is usually at a low pace and the variance remains significant even for large k* (*see Fig.* 2).[10]

Hypothesis 3.1 The RL standard deviation is a non-decreasing function of a with a possible decrease for larger values of a. The general form is an increasing or inverse U-shaped function.

Hypothesis 3.2 The RL standard deviation is a non-decreasing function of k for low values of k, with the possibility of a decrease for larger values of k.

[10] Result (ii) (variation with respect to k) is not proven analytically but only stands as a numerical result.

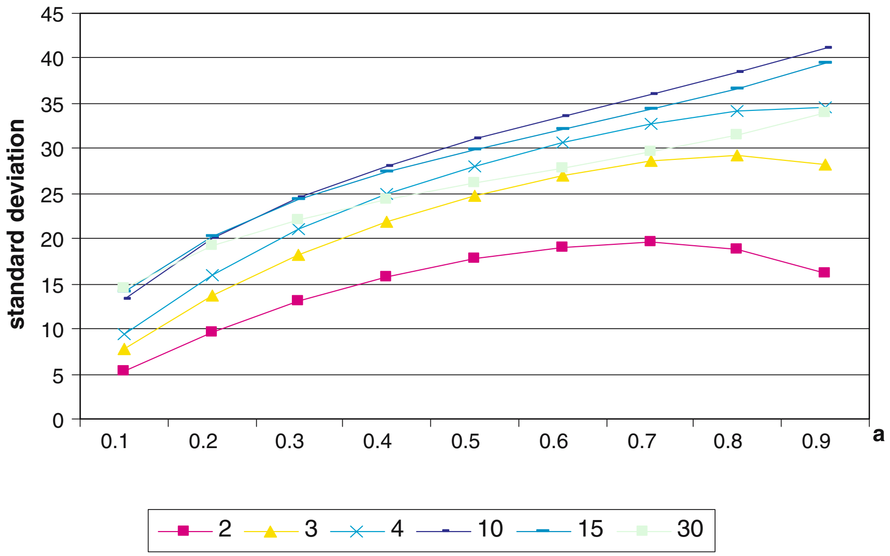

Fig. 1 Price dispersion as a function of a (for different values of k)

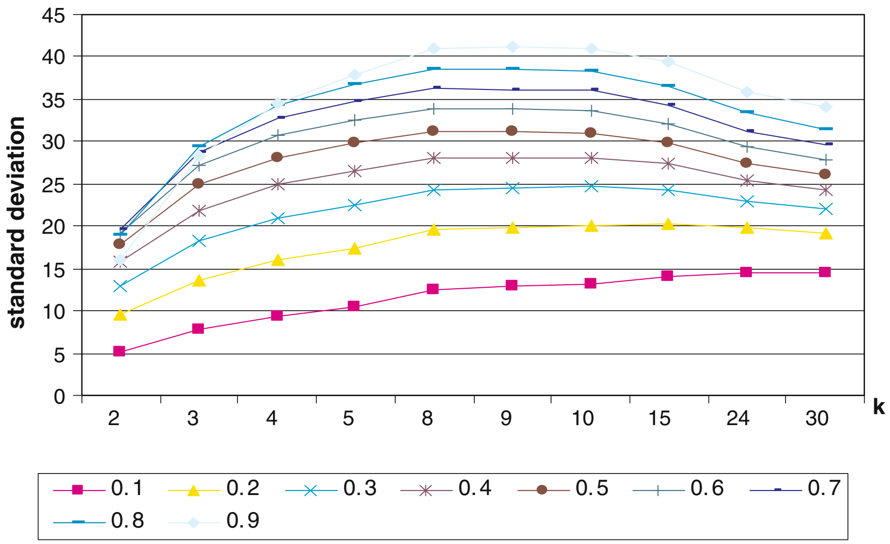

Fig. 2 Price dispersion as a function of k (for different proportions of informed consumers)

3.2 Test implementation

To compare NSE with RL outcomes, for each parameters configuration (a, k) we ran 25 sessions.[11] We define each simulation as a session. For each session,

[11] The process has been simulated in Java: the appendix reports an abridged pseudo code of the model. Source code, classes, pseudo code and an executable version of the program are available at http://e.darmon.free.fr/fssmarket/ or by request to the authors.

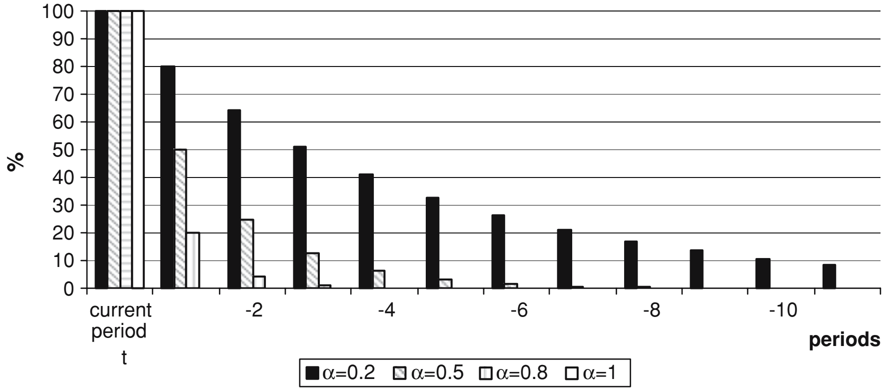

Fig. 3 Weighting of profits in the fitness function at time $t+1$ (in% of current weight α)

we computed the posted price distribution, the aggregate mean and the standard deviation of the posted price distribution over the last 100 periods (see the Appendix for the definition of the statistics).

We present the results of simulations on (a, k) using a model where temperature (the exploration/exploitation coefficient) is $\tau = 0.05$, the reward updating coefficient $\alpha = 0.8$, the initial fitness coefficient $\delta = 0.2$ and the number of rounds is $T = 1{,}000$. The number of rounds (T) has been chosen so as to ensure that the process converged to a stationary position within this time horizon. Thus, as a benchmark, we consider a situation where (i) individuals' initial expectations are not too optimistic ($\delta = 0.2$), (ii) price volatility is not too high or random (as it might be with higher temperature)[12] and (iii) sellers' memory is not too long ($\alpha = 0.8$). Typically, for $\alpha = 0.8$, only profits over the two last recorded profits conserve a significant weight compared to the weight of current profit in the fitness of the rule (see Fig. 3). When $\alpha = 1$, only the current profit is taken into account in the fitness function.

Let us refer to this default parameter configuration as the *reference configuration*. We first test the previous hypotheses in the reference configuration: to test H0, we implemented a Kolmogorov–Smirnoff adequacy test (testing that the RL and the NSE distributions are equal). This test holds only for continuous distribution i.e., for any value of (a, k) such as $a \neq 0$ and $a \neq 1$ and $k \neq 1$. The case $a = 1$ will be detailed hereafter. For the remaining hypothesis (H1.1 to H3.2), one can note that the form of the tested hypothesis is weaker than its related NSE proposition. Consider e.g., Proposition 1 and Hypothesis 1.2. We test that the average price is "non-decreasing" while the related Proposition states that the average price is increasing with k. Testing the assertion that the

[12] The choice of the reference configuration is somehow arbitrary. However, temperature (whenever not too large) and initial fitness seems to have no effect on the results of comparative statics performed on (a, k). Temperature adds noise to the system while initial fitness seems to speed up the learning convergence. For the reward updating coefficient α, assuming a not too long memory is reasonable given our limited rationality assumption.

average price is “non-decreasing” in k rather than “increasing” comes to the test that the average price is either (i) significantly increasing, (ii) non-significantly increasing or (iii) non-significantly decreasing while testing the increase of the average price would require that (ii) or (iii) be false.

There are two reasons for choosing this weaker form. First, unlike analytical outcomes, RL-outcomes necessarily exhibit some statistical randomness. This is partly due to the persistence of explorative behaviors (introducing some noise to the observations) even when convergence is achieved. Since a “nondecrease” test is less demanding that an “increase” test, we take as being true, cases which otherwise would have been rejected by the “increasing” test although they may have been true without statistical randomness. In addition, some variations observed at NSE may be quite small and fall within some measure of statistical randomness. For example the increase in the Nash average price from $k = 11$ to $k = 12$ and for $a = 0.1$ is equal to 0.19. This is clearly below the estimate noise of 0.287 generated by the process (see below). Second, from a more methodological point of view, we use a falsification criteria i.e., we do not necessarily want Nash predictions to be true but at least we want them not to be false.

To evaluate the amount of statistical randomness, we simply take the maximal difference obtained from parameter configurations $a = 0$ or $k = 1$. Indeed, when $a = 0$ or $k = 1$, all consumers are uninformed and actual differences in the mean price can be directly attributed to statistical randomness. To illustrate this, let us consider the case of the average price. The following tables report the observed average prices as $a = 0$ or $k = 1$ (for the reference configuration):

k=	1	2	3	4	5	6	7	8	9	10
a=0	79.89	79.77	79.84	79.78	79.76	79.75	79.76	79.82	79.74	79.86

k=	11	12	13	14	15	16	17	18	19	20
a=0	79.82	79.71	79.75	79.89	79.91	79.84	79.92	79.82	79.68	79.82

a=	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
k=1	79.89	79.88	79.75	79.77	79.63	79.80	79.74	79.83	79.76	79.81

The maximum difference is here equal to 0.287 (difference between average prices of ($a = 0, k = 17$) and ($a = 0.4, k = 1$). We define this as a measure of the randomness when sampling in the reference configuration and define it as the threshold above which one can conclude that a series is increasing (or decreasing).

Finally, each test implemented in the reference configuration, is complemented by similar tests performed with the other parameter configurations (Table 1). This enables us to assess the robustness of the conclusion obtained in the reference configuration. We considered these other parameters configurations. Except for configuration 9 and 10, these are just one-dimension variations in the parameter space $\{T, \alpha, \delta, \tau\}$ of the reference configuration $\{T = 1{,}000; \alpha = 0.8; \delta = 0.2\}$ (configuration 1).

Table 1 List of all parameter configurations tested

	α	T	F_0	Temperature
Configuration 1	0.8	1,000	0.2	0.05
Configuration 2	0.8	10,000	0.2	0.05
Configuration 3	0.4	1,000	0.2	0.05
Configuration 4	0.4	10,000	0.2	0.05
Configuration 5	1	1,000	0.2	0.05
Configuration 6	0.8	1,000	0.6	0.05
Configuration 7	0.8	1,000	1	0.05
Configuration 8	0.8	1,000	0.2	0.1
Configuration 9	0.3	2,500	0.6	0.05
Configuration 10	0.5	1,000	0.6	0.05

4 Results

4.0 Price distribution

Result 0 *The stationary distribution with reinforcement learners does not converge to the Nash distribution for* $a \neq 1$.

With a 5% significance level, adequacy tests systemically reject the assumption that the RL- and the NSE-distributions are equal. This holds for all parameter (a, k) such that $0 < a < 1$ and $k \neq 1$. Besides, it has been first obtained with the reference configuration but has been also extended to the other parameter configurations. This result has been tested for each individual session in all configurations (i.e., for 10×25 sessions) and is in line with some theoretical results in the literature concerned with learning in games. Hopkins and Seymour (2002) show that Varian's model is unstable under a variety of learning models including reinforcement learning except if there are sufficiently uninformed consumers.[13] One can thus wonder why we do not observe this convergence. One reason might be the approximation of a continuous NSE by a discrete support. As outlined by these authors, if the discrete approximation of the continuous distribution is not precise enough, this approximation is not necessarily a fixed point of the continuous positive definite adaptive dynamics used to prove their result. However, taking a finer price grid has two drawbacks. First, consumers may not be sensitive to a reduction of some cents in prices. Second, sellers are able only to learn about a finite and reasonable number of prices. But another reason is that for $\tau > 0$, the dynamics is not a positive definite adaptive one.[14]

In the case $a = 1$, convergence to Bertrand equilibrium mainly depends on the values of the reward updating coefficient α, T (number of time periods) and the number of visits made by informed consumers k. In the following figures,

[13] Specifically when informed consumers visit S sellers, the condition is $1 - a > \frac{S}{2(S-1)}$.

[14] In fact, given the logit learning rule, the fixed points are not Nash and convergence should approach a logit equilibrium when α is small (see for example, Capra et al. (2002)).

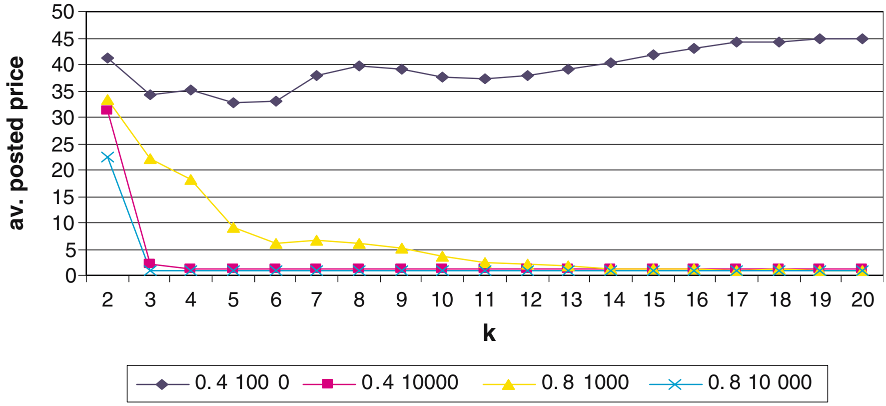

Fig. 4 Average posted price, $a = 1$

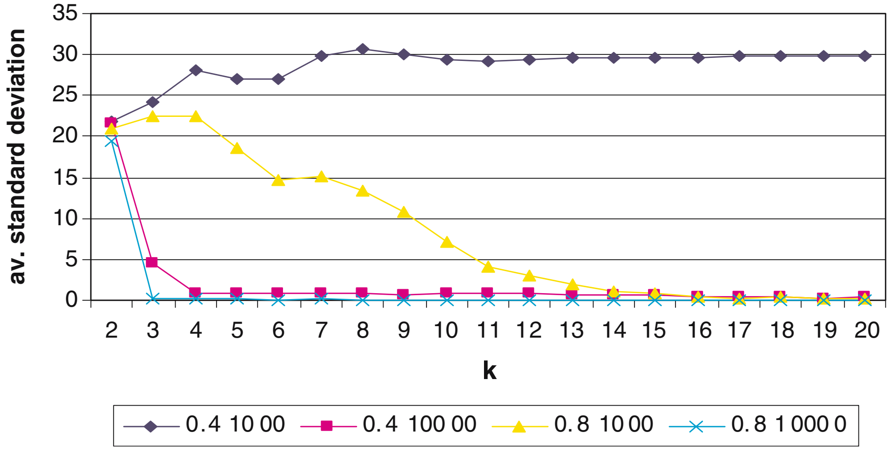

Fig. 5 Average standard deviation posted

we display the average posted price and standard deviation for α set to 0.4 or to 0.8, and T (number of time periods) set to 1,000 or 10,000 (*ceteris paribus*). Figures 4 and 5 show the average of these statistics over 25 sessions for all parameters configurations.

The convergence of the average price to the Nash equilibrium depends on the reward updating coefficient α. For $\alpha = 0.4$ and $T = 1,000$, average prices did not converge to the Nash equilibrium but raising the number of time periods to 10,000 was enough to restore it. In addition, the number of visits made by informed consumers k affects the convergence to the Bertrand equilibrium. With $k = 2$ average posted price where clearly different from Nash. In the case where $\alpha = 0.8$ and $T = 1,000$ convergence was only for values of $k > 13$. However, with $\alpha = 0.8$ and $T = 10,000$, this was no longer the case. It is thus interesting to note that the only case exhibiting convergence to the Nash equilibrium is the case where $a = 1$ (no uninformed buyers) i.e., where competition

Table 2 Theoretical mean market price versus RL mean price for $v = 100$

	k													
a	2		3		4		5		8		10		15	
0	*100*	**79.8**	*100*	**79.8**	*100*	**79.8**	*100*	**79.8**	*100*	**79.8**	*100*	**79.9**	*100*	**79.9**
0.1	*90*	**75.7**	*91*	**75.3**	*91*	**75.3**	*91*	**75.9**	*92*	**77.2**	*93*	**77.8**	*94*	**79.0**
0.2	*81*	**71.5**	*82*	**70.7**	*84*	**71.4**	*85*	**72.2**	*87*	**74.7**	*88*	**75.9**	*90*	**77.6**
0.3	*72*	**67.2**	*75*	**66.3**	*77*	**67.4**	*79*	**68.8**	*82*	**72.1**	*84*	**73.7**	*87*	**76.5**
0.4	*64*	**62.4**	*68*	**61.7**	*71*	**63.2**	*73*	**65.0**	*79*	**69.5**	*81*	**71.6**	*85*	**75.3**
0.5	*55*	57.5	*60*	**56.9**	*65*	**58.9**	*68*	**61.3**	*75*	**67.0**	*78*	**69.5**	*83*	**74.2**
0.6	*46*	52.3	*53*	**51.7**	*59*	**54.4**	*63*	**57.3**	*71*	**64.1**	*75*	**67.4**	*80*	**72.9**
0.7	*37*	47.0	*46*	**45.9**	*52*	**49.5**	*57*	**52.9**	*67*	**61.1**	*71*	**64.9**	*78*	**71.4**
0.8	*27*	41.7	*37*	39.5	*45*	**43.5**	*51*	**47.6**	*63*	**56.9**	*67*	**61.1**	*75*	**68.9**
0.9	*16*	37.0	*27*	31.5	*35*	35.7	*42*	**40.0**	*56*	**50.2**	*62*	**55.0**	*71*	**63.5**
1.0	*1*	33.3	*1*	22.1	*1*	18.3	*1*	9.2	*1*	6.2	*1*	3.6	*1*	1.2

Note: The values in italic (resp. normal values) depict the NSE predictions (resp. the RL outcomes) of the average posted price (average computed over all prices set over the last 100 periods and ran for 100 sessions in the reference case). The bold figures indicate that RL mean price is lower than the NSE mean price

is à *la* Bertrand[15] but that this convergence may depend on the time scale T, the behavior of informed consumers (k) and the memory of sellers' learning (α).

4.1 Mean posted price

Table 2 sketches the average price of the RL distribution and compares it to the NSE expected price.

Result 1.1 *The RL mean price is a decreasing*[16] *function of a* (*proportion of informed consumers in the market*) *and an increasing function of k (except for* $k = 2$*) as shown in Table* 2.

To test for statistical significance, we used the procedure described in the previous section: as indicated, two mean prices that are generated by different values of a or k, are statistically different if the difference between the two values exceeds the threshold measuring statistical randomness. It has been evaluated for the mean posted price in the reference configuration to 0.287. Whenever the (absolute) variation between two prices is greater than the threshold, we consider that this variation is significant. Otherwise, the variation is not significant i.e., meaning that average prices are stable.

The NSE mean price is in general higher than the RL mean price except when all consumers are informed ($a = 1$) or for values of k less than or equal to 4 and for a large proportion of informed consumers.

To check whether Hypotheses 1.1 and 1.2 hold, the parameter configurations of table 1 were tested.

15 See also Darmon and Waldeck (2005) for a graphical presentation of the price distribution in this case $a = 1$ and for a particular general case $a \neq 1$.

16 By decreasing we do not mean here significantly decreasing but just a "rough" comparison of average prices.

Table 3 Test of hypothesis H 1.2 for configurations 1 to 10 for $k = 2$ (for all k#2, all other figures = 1 expect when $k = 3$ and $a = 0.9$ for configuration 3 and 10)

				a								
RUP	T	F_0	Temperature	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.8	1,000	0.2	0.05	1	0	0	0	0	0	0	0	0
0.8	10,000	0.2	0.05	0	0	0	0	0	1	0	0	0
0.4	1,000	0.2	0.05	1	0	0	0	0	0	0	0	0
0.4	10,000	0.2	0.05	1	1	1	1	1	1	0	0	0
1	1,000	0.2	0.05	0	0	0	0	0	1	1	0	0
0.8	1,000	0.6	0.05	0	0	0	0	0	0	0	0	0
0.8	1,000	1	0.05	1	0	0	0	0	0	0	0	0
0.8	1,000	0.2	1	1	0	0	0	0	0	0	0	0
0.3	2,500	0.6	0.05	1	1	1	1	1	1	0	0	0
0.5	1,000	0.6	0.05	1	1	1	0	0	0	0	0	0

Note: Digit '1' indicates that the hypothesis is true and digit '0' that it is false. Since with H1.2, we test the non-decreasing nature of the average posted price with respect to k: a '0' digit for $k = 2$ indicates that the average posted price is significantly decreasing from $k = 2$ to $k = 3$

Result 1.2 (generalization) *With very few exceptions,*[17] *Hypothesis* 1.1 *is true for all parameter configurations that is the RL mean price is non-increasing in a. Hypothesis* 1.2 (*i.e., the RL mean price is non-decreasing in k*) *is true with the exception of some cases when* $k = 2$ *as indicated by Table* 3 *and two cases* ($k = 3$ *and* $a = 0.9$ *for configuration* 3 *and* 10).

We performed a one sided Wilcoxon ranked test (at 5%).[18] For this test, 98.2% of the figures exhibited a variation in line with Nash, meaning that 98% of cases showed a significant decrease in average price when a increases from a to $a + 0.1$. None of the cases showed however a significant increase. This is conform to result 1.2. The same test for k indicates that 82.6% of the variations were conform to NSE (increasing in k) whereas only 5.6% of all figures where in contradiction to NSE prediction. Moreover, these figures were for $k = 2$ only in accordance with Table 3 for configuration 1.

4.2 Average price paid by informed buyers

Table 4 presents the RL mean price paid by informed buyers for different values of a and k. We can observe that the average price paid by informed consumers at NSE is larger than the average price in RL for low values of a.

17 The two only exceptions are for: (α, T, F_0, Temperature, k) = (0.8, 10,000, 0.2, 0.05, 20) for a increasing from 0.6 to 0.7 and (α, T, F_0, Temperature, k) = (0.4, 10,000, 0.2, 0.05, 20) for a increasing from 0.8 to 0.9 for which average prices increase significantly, that is the increase exceeds the estimated noise.

18 For the reference configuration, we tested whether over the 25 sessions there is a significant number of cases where the average price decreased (resp. increased) with an increase in a (resp. k) by steps of 0.1 (resp. 1).

Table 4 Average accepted price (informed buyers)

	k											
a	2		3		5		8		10		15	
0.1	*87*	63.4	*84*	54.0	*78*	43.0	*70*	33.3	*66*	28.7	*58*	21.8
0.2	*76*	59.1	*70*	49.5	*62*	39.1	*53*	30.8	*48*	27.0	*40*	20.2
0.3	*65*	54.8	*59*	44.9	*50*	35.3	*41*	27.7	*37*	24.3	*30*	18.5
0.4	*55*	50.0	*49*	40.1	*40*	31.1	*32*	24.4	*29*	21.3	*23*	16.5
0.5	*45*	45.1	*40*	35.2	*32*	26.8	*25*	21.0	*22*	18.5	*17*	14.6
0.6	*36*	40.1	*31*	29.9	*25*	22.3	*19*	17.4	*17*	15.3	*13*	12.2
0.7	*27*	34.9	*23*	24.3	*18*	17.8	*14*	13.8	*12*	12.2	*9.4*	9.8
0.8	*18*	29.7	*16*	18.5	*12*	13.0	*9.4*	9.8	*8.1*	8.7	*6.2*	7.1
0.9	*9.3*	25.3	*8.2*	12.3	*6.4*	7.9	*4.9*	5.7	*4.2*	5.0	*3.2*	3.6
1	*1*	21.7	*1*	6.7	*1*	1.0	*1*	1.0	*1*	1.0	*1*	1.0

Note: The values in italic (resp. normal values) depict the NSE prediction (resp. the RL outcome) of the average price accepted by informed buyers (reference configuration)

Result 2.1 *The RL mean price paid by informed consumers decreases significantly with a for all values of* (a, k) *except one case as shown in Table* 5 *for the reference configuration.*[19]

A Wilcoxon ranked test leads to the same conclusion: only one case shows a non-significant decrease but the increase is not significant for this case.

Result 2.2 *The RL mean price paid by informed consumers decreases significantly with k, except for some values of a and k as shown in Table* 5 *for the reference case* (27% *of the cases*). *However for these values the increase is not significant.*

A Wilcoxon ranked test leads to the same conclusion: 80.7% of the figures report a significant decrease in k but none has a significant increase.

Again, we iterated the same tests for alternative parameter configurations.

Result 2.3 (generalization) *Hypotheses H2.1 and H2.2 are true for all parameter configurations* 1 *to* 10*, that is the average price paid by informed consumers never increases significantly with a or k.*[20]

4.3 Standard deviation of posted prices

In the reference configuration, Fig. 6 displays the variations of the NSE- and RL-standard deviation of posted prices as a function of a, for different scenarios of k $(k = 2, 5, 10, 15, 18)$.

[19] We have the stronger result of a significant decrease rather than a non-increase.

[20] For k, three figures did not satisfy H2.2 for the configuration $(\alpha; T; F_0; \tau) = (0.8; 10{,}000; 0.2; 0.05)$ and two figures for $(\alpha; T; F_0; \tau) = (1; 1{,}000; 0.2; 0.05)$.

Table 5 Average accepted price for informed consumers with significance tests

	a									
k	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
2	63;1;1	59;1;1	55;1;1	50;1;1	45;1;1	40;1;1	35;1;1	30;1;1	25;–;1	22;–;–
3	54;1;1	49;1;1	45;1;1	40;1;1	35;1;1	30;1;1	24;1;1	19;1;1	12;–;1	7;–;–
4	47;1;1	44;1;1	39;1;1	35;1;1	30;1;1	25;1;1	20;1;1	15;1;1	9;–;1	2;–;–
5	43;1;1	39;1;1	35;1;1	31;1;1	27;1;1	22;1;1	18;1;1	13;1;1	8;–;1	1;–;–
6	39;1;1	35;1;1	32;1;1	28;1;1	24;1;1	20;1;1	16;1;1	12;1;1	7;–;0	1;–;–
7	36;1;1	32;1;1	30;1;1	26;1;1	22;1;1	19;1;1	15;1;1	11;1;1	6;–;0	1;–;–
8	34;1;1	31;1;1	28;1;1	24;1;1	21;1;1	17;1;1	14;1;0	10;1;0	6;–;0	1;–;–
9	30;1;1	28;1;1	26;1;1	23;1;1	19;1;1	16;1;1	13;1;1	9;1;0	6;–;0	1;–;–
10	29;1;1	27;1;1	25;1;1	22;1;1	19;1;1	15;1;1	12;1;1	9;1;1	5;–;0	1;–;–
11	27;1;1	25;1;1	22;1;1	20;1;1	17;1;1	14;1;0	11;1;0	8;1;1	5;–;1	1;–;–
12	26;1;1	23;1;1	21;1;1	19;1;1	17;1;1	14;1;0	11;1;1	8;1;0	4;–;0	1;–;–
13	24;1;1	22;1;1	20;1;0	18;1;1	16;1;0	13;1;1	10;1;0	8;1;1	4;–;1	1;–;–
14	22;1;1	21;1;1	20;1;1	17;1;1	15;1;1	13;1;0	10;1;0	7;1;0	4;–;0	1;–;–
15	22;1;1	20;1;1	18;1;1	16;1;1	15;1;0	12;1;0	10;1;0	7;1;0	4;–;0	1;–;–
16	20;1;0	19;1;1	17;1;0	16;1;1	14;1;1	12;1;1	10;1;0	7;1;0	3;–;0	1;–;–
17	20;1;1	18;1;1	17;1;1	15;1;1	13;1;1	11;1;0	9;1;1	7;1;0	3;–;0	1;–;–
18	18;1;0	17;1;0	16;1;0	14;1;0	13;1;0	11;1;0	8;1;0	7;1;0	3;–;0	1;–;–
19	18;1;0	17;1;1	15;1;1	14;1;1	12;1;1	10;1;1	8;1;1	7;1;1	3;–;0	1;–;–
20	18;1;–	16;1;–	14;1;–	13;1;–	11;1;–	9;1;–	7;0;–	6;1;–	3;–;–	1;–;–

Note: Each cell has three components: the first number indicates the average accepted price by informed consumers. The second component reports the result of the test that average accepted price by informed consumers decreases with a (horizontal reading). The third one reports the result of the test that the average accepted price decreases with k (vertical reading). Digit '1' (resp. '0') indicates that the test is true (resp. false). Mark "–" indicates that the test is not relevant for these values of the (a, k) couple (border values). The statistical randomness for this table amounts to 0.594. For example, for cell $(a, k) = (0.8, 15)$ we have numbers 7; 1; 0 meaning that (1) average accepted price by informed consumers is 7.1; (2) the decrease in average accepted price from $a = 0.8$ to $a = 0.9$ is significant (7.1–3.6 > 0.594) and (3) that the decrease from $k = 15$ to $k = 16$ is not significant $|7.1 - 7.3| < 0.594$

From the Fig. 6, we can infer that RL-standard deviation is decreasing for $k = 2$. For the other values of k, the RL standard deviation is an inverse U-shaped function of a with a maximum at $a = 0.8$ for $k = 3$ and 4 (Table 6 for the reference configuration). For larger values of k, the maximum is at (or above) $a = 0.9$. Besides, we can notice that although price dispersion increases as a function of a (except when $k = 2$), the variations are less pronounced for RL than for NSE.

The same remark holds for variations with respect to k as pictured by Fig. 7. Similarly, this figure displays the variations of the NSE- and RL-standard deviation of posted prices as a function of k, for different scenarios of a ($a = 0.1, 0.3, 0.7, 0.9$) in the reference configuration.

From Fig. 7 and Table 6, we can infer the variations of the standard deviation with respect to k are inverse U shaped function for both distributions (RL and NSE) except for the RL distribution when $a = 0.1$. Apart this exception, for both distribution, the variance is steadily increasing for low values of k, whereas it decreases at a lower pace for larger values of k.

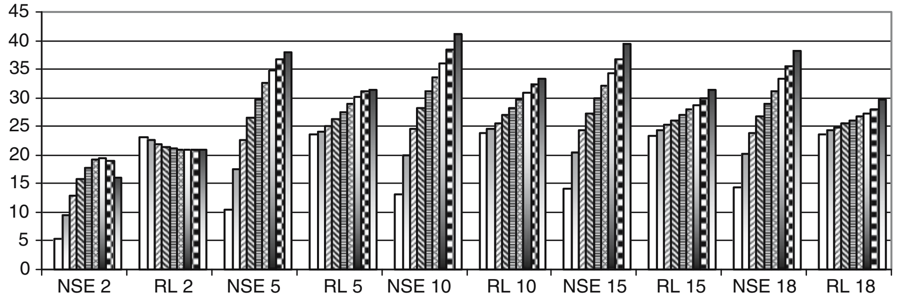

Fig. 6 Standard deviation of price as a function of k (comparison of NSE to RL for different scenarii of $k = 2, 5, 10, 15, 18$)

Table 6 Standard deviation of posted price of RL distribution

	k												
a	2	3	4	5	6	7	8	9	10	12	15	18	20
0	23.5	23.4	23.4	23.4	23.4	23.4	23.3	23.3	23.4	23.5	23.3	23.3	23.3
01.1	23.2	23.4	23.7	23.7	23.7	*23.8*	23.7	23.7	23.8	23.7	23.5	23.5	23.5
0.2	22.6	23.3	23.9	24.2	24.5	*24.6*	24.5	24.6	24.5	24.5	24.3	24.2	24.2
0.3	21.9	23.4	24.4	25.0	25.4	*25.7*	25.6	25.7	25.6	25.5	25.2	24.9	24.7
0.4	21.4	23.7	25.3	26.2	26.8	27.0	*27.1*	27.0	27.0	26.7	26.1	25.6	25.2
0.5	21.1	24.3	26.5	27.6	28.1	*28.4*	28.4	28.4	28.3	27.9	27.0	26.1	25.5
0.6	20.9	25.1	27.6	28.9	29.5	*29.9*	29.8	29.8	29.6	29.0	27.9	26.7	25.7
0.7	20.8	25.7	28.6	30.1	30.8	*31.2*	31.2	31.1	30.9	30.1	28.8	27.3	25.9
0.8	20.8	**26.0**	**29.3**	31.1	32.0	32.5	32.6	32.5	32.2	31.5	29.8	27.9	25.9
0.9	20.8	25.3	29.3	**31.3**	**32.7**	**33.4**	***33.7***	**33.4**	**33.4**	**32.8**	**31.3**	**29.6**	**27.5**
1	21.0	22.5	22.4	18.7	14.7	15.2	13.5	10.9	7.2	3.1	0.8	0.5	0.3

Note: The table displays the standard deviation (averaged over 25 sessions) in the reference configuration for the RL distribution. The bold numbers represent the maximum standard deviation for a given k. The values in italic and bold italic represent the maximum standard deviation for a given a

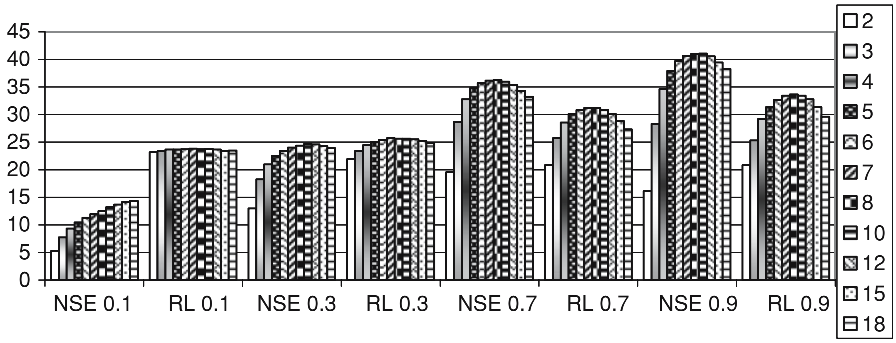

Fig. 7 Standard deviation of price as a function of k (compared NSE and RL for different scenarii of $a = 0.1; 0.3; 0.7; 0.9$)

Table 7 Test 3.1: standard deviation as a function of *a*

	a							
k	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
2	00	00	00	00	00	01	01	01
3	01	01	11	11	11	11	11	00
4	01	11	11	11	11	11	11	01
5	11	11	11	11	11	11	11	11
6	11	11	11	11	11	11	11	11
7	11	11	11	11	11	11	11	11
8	11	11	11	11	11	11	11	11
9	11	11	11	11	11	11	11	11
10	11	11	11	11	11	11	11	11
11	11	11	11	11	11	11	11	11
12	11	11	11	11	11	11	11	11
13	11	11	11	11	11	11	11	11
14	11	11	11	11	11	11	11	11
15	11	11	11	11	11	11	11	11
16	11	11	11	11	11	11	11	11
17	11	11	11	11	11	11	11	11
18	11	11	11	11	11	11	11	11
19	11	11	11	11	11	11	11	11
20	11	11	11	11	01	01	01	11

Let us test *Hypothesis* 3.1 that the RL standard deviation is a non decreasing function of *a* with the possibility of a decrease for larger values of *a* that is the general form is an increasing function or inverse U-shaped. For that, we implemented two parallel tests: first, we tested for a significant *increase* in standard deviation with respect to *a*. Second, we tested for a *non-decrease* in standard deviation. Table 7 exhibits the result of this test for the reference configuration: in each cell, there are two digits. The first one refers to the test for a significant *increase*: thus '1' means that the increase in standard deviation is significant and '0' that we have either (i) a significant decrease, (ii) a non-significant decrease or (iii) a non-significant increase. The second digit refers to the test of a non-significant decrease: thus digit '1' means either a significant increase, a non-significant decrease or a non-significant increase while '0' means a significant decrease. With these two pieces of information, each cell may be interpreted as follows: string '11' means a significant increase in standard deviation, '10' is a impossible case, '01' means a stable situation with either a non-significant decrease or a non-significant increase, and finally '00' means a significant decrease.

Testing for Hypothesis 3.1 means that the following sequence of strings is not allowed: for each row in Table 7, once a '00' cell appears (a significant decrease), it is never followed (in an horizontal reading) by a '11' cell (significant increase). Similarly, testing Hypothesis 2.1 means that the following sequence of strings is not allowed: for each column, in Table 8, a '00' cell is never followed by a '11' cell (vertical reading).

Table 8 Test 3.2: standard deviation as a function of k

	a								
k	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	01	11	11	11	11	11	11	11	11
3	11	11	11	11	11	11	11	11	11
4	01	11	11	11	11	11	11	11	11
5	01	11	11	11	11	11	11	11	11
6	01	01	11	01	11	11	11	11	11
7	01	01	01	01	01	01	01	01	11
8	01	01	01	01	01	01	01	01	00
9	01	01	01	01	01	01	01	00	01
10	01	01	01	01	01	00	00	00	00
11	01	01	01	01	01	00	00	00	00
12	01	01	01	00	00	00	00	00	00
13	01	01	00	01	00	00	00	00	00
14	01	01	01	01	00	00	00	00	00
15	01	01	01	01	00	00	00	00	00
16	01	01	01	01	00	00	00	00	00
17	01	01	01	01	00	00	00	00	00
18	01	01	01	00	00	00	00	00	00
19	01	01	01	01	01	00	00	00	00

Result 3.1 *H3.1 is true for the reference configuration except when $k = 2$ where the decrease in variance is significant with a* (*cf.* Table 7). *In general, standard deviation is an increasing function of a. For the case $k = 3$, we observe a significant decrease for a = 0.8 to a= 0.9. This is in line with the Nash result .*

In addition, we performed a one sided Wilcoxon ranked test (at 5%) for the variation with respect to a in the reference case. 83% of figures reported a variation conform to the NSE prediction whereas 3% showed a opposite variation to NSE. The figures showing a variation contradicting the NSE prediction for the Wilcoxon test were solely for $k = 2$.

Result 3.2 *H3.2 is true for the reference configuration* (*cf.* Table 8). *In general, standard deviation is significantly increasing up to $k = 6$ and remains stable or decreases afterwards. One exception is for $a = 0.1$ for which standard deviation is rather stable.*

Result 3.3 (generalization) *Result* 3.1 *holds for parameter configurations* 1 *to* 10 *i.e., H3.1 is true except when $k=2$.*[21] *Result* 3.2 *also holds for parameter configurations* 1 *to* 10 *i.e., H3.2 is true. Moreover, in general standard deviation is significantly increasing up to $k = 5$ or 6, except for the case $a = 0.1$ for which standard deviation is stable with respect to k.*

Table 9 shows some other parameter configurations:

[21] There are also two more minor exceptions: in Configuration 2, for $k = 20$, a decrease for $a = 0.8$ is followed by an increase for $a = 0.9$. In Configuration 3 and $k = 3$, the variance is significantly decreasing for $a = 0.1$ and remains stable afterwards.

Table 9 Standard deviation with respect to k for different parameter configurations (α, T, F_0, τ)

0,8 10,000, 0.2, 0.05									
a	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
k									
2	01	01	11	11	11	11	11	11	11
3	01	11	11	11	11	11	11	11	11
4	01	11	11	11	11	11	11	11	11
5	01	01	01	11	11	11	11	11	11
6	01	01	01	01	01	01	01	01	01
7	01	01	01	01	01	01	01	01	11
8	01	01	01	01	01	01	01	01	01
9	01	01	01	01	01	01	01	01	01
10	01	01	01	01	01	01	01	00	00
11	01	01	01	01	01	01	00	01	00
12	01	01	01	01	01	01	00	01	01
13	01	01	01	01	00	01	01	00	00
14	01	01	01	01	01	01	01	00	01
15	01	01	01	01	01	00	00	00	00
16	01	01	01	01	01	01	00	00	00
17	01	01	01	01	00	00	00	00	00
18	01	01	01	01	01	01	00	00	01
19	01	01	01	01	01	00	00	00	00

0.4, 1,000, 0.2, 0.05									
a	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
k									
2	01	11	11	11	11	11	11	11	11
3	01	11	11	11	11	11	11	11	11
4	01	11	11	11	11	11	11	11	11
5	01	01	11	11	11	11	11	11	11
6	01	11	01	01	11	11	11	11	11
7	01	01	01	11	01	01	01	11	11
8	01	01	01	01	01	01	01	01	11
9	01	01	01	01	01	01	01	01	01
10	01	01	01	01	01	01	01	01	01
11	01	01	01	01	01	01	01	01	01
12	01	01	01	01	01	01	01	01	01
13	01	01	01	01	01	01	01	01	01
14	01	01	01	01	01	01	01	01	01
15	00	01	01	01	01	01	01	01	01
16	01	01	01	01	01	01	01	01	01
17	01	01	01	01	01	01	01	01	01
18	01	01	01	01	01	01	01	01	01
19	01	01	01	01	01	01	01	01	01

0.4, 10,000, 0.2, 0.05									
a	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
k									
2	01	11	11	11	11	11	11	11	11
3	01	11	11	11	11	11	11	11	11
4	01	01	11	11	11	11	11	11	11
5	01	01	11	11	11	11	11	11	11
6	01	01	01	01	01	11	11	11	11
7	01	01	01	01	01	01	01	01	01
8	01	01	01	01	01	01	01	01	01
9	01	01	01	01	01	01	01	01	01
10	01	01	01	01	01	01	01	01	01
11	01	01	01	01	01	01	01	01	01
12	01	01	01	01	01	01	01	01	01
13	01	01	01	01	01	01	01	01	01
14	01	01	01	01	01	01	00	01	00
15	01	01	01	01	01	01	01	00	00
16	01	01	01	01	01	01	01	00	00
17	01	01	01	01	01	01	00	00	00
18	01	01	01	01	01	01	01	00	00
19	01	01	01	01	01	01	01	00	00

5 Conclusion and future work

In this paper, we established two results. First, we showed that in the general case (equilibrium in mixed strategies), sellers using reinforcement learning do not learn to play the Nash equilibrium price strategy. Second, the original contribution of this paper is to show that, even if the dynamics does not converge to Nash, Nash predictions are performing remarkably well with respect to a change in the degree of information (parameters a and k). In other words, we showed that Nash predictions under individual learning are still valid when these two parameters vary: as predicted by Nash, the mean posted price is non-increasing with a and non-decreasing with k (except for $k = 2$). The mean price accepted by informed buyers in non-decreasing with both a and k. Except for $k = 2$, and in accordance with Nash, price variance is in general increasing with a. Finally, price dispersion is an inverse U-shaped function of k except for the case $a = 0.1$ for which it is rather stable. These results have been obtained with a "reference" parameter configuration but we tested alternative configurations and found that these conclusions are robust to changes in parameters specification. Especially, these conclusions are not sensitive to a change in RL-parameters.

One interesting point is now to link our results to the observations made in the lab. In the context of a search market, Morgan et al. (2006) set up an experimentation where they compare four treatments: $a = 1/2$ or 5/6 and $S = k = 2$ or $S = k = 4$. Their observations corroborate both the Nash and the reinforcement learning results. That is, the average posted price is decreasing in a and increasing in $k = S$, the average price paid by informed consumers is decreasing in a and $k = S$, and price dispersion is increasing in a and $k = S$ as predicted within the parameter range tested. However, in view of our results, the experiment which was supposed to be a confirmation in favor of Nash assumption, may also be a confirmation for the RL hypothesis. The appropriate behavioral model for this kind of market still remains an open question.

Other issues are also open. First, our analysis was on the distribution of prices at the market-level, and not at the firm level. Yet, a question is whether firms learn to fix the same average price. To test this assumption, we conducted a Fisher-Snedecor test for the 100 sessions of the reference case. Let H0 be the assumption that all firms set an equal average price over the last 100 periods. Table 10 shows that in general firms learn different prices. This is especially true when a is greater than 0.4 and $k > 3$ for which in general at least one firm fixes a different mean price from the rivals. For example for $k = 20$ and $a = 1$ H0 was true for only three sessions out of 100.

The exact form of individual learning developed by sellers is still to be explored. Especially, the market game proposed may typically exhibit some Edgeworth cycle. As already noted, one could also consider alternative types of learning, including EWA (Camerer and HO 1999) or social learning. Finally, the market game may be extended in a few directions. One is by introducing endogenous search. Another is to introduce learning on buyers' side. In a richer framework, price dispersion in a market can be comprehended as the consequence of a *self-organization* process through adaptive learning and

Table 10 Percentage of time H0 is true at 5% level

	k																		
a	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0.0	66	67	57	52	62	59	57	57	53	64	55	62	60	53	53	61	65	59	63
0.1	69	51	51	41	47	43	38	37	51	43	45	42	42	44	50	40	54	45	43
0.2	64	38	33	28	17	21	18	26	24	22	29	27	21	19	25	22	23	18	19
0.3	51	28	10	11	7	10	7	8	7	6	5	8	4	13	6	8	9	11	8
0.4	45	15	4	3	8	1	4	4	0	3	5	1	1	4	3	1	2	1	2
0.5	33	13	3	2	0	0	0	0	1	0	0	0	1	0	2	0	0	0	0
0.6	37	8	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.7	33	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.8	38	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.9	47	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1.0	45	34	21	6	4	2	0	0	0	10	10	24	14	17	10	13	11	11	3

co-evolution by all agents as shown for example in Kirman and Vriend (2001). Thus, buyers' search behavior could integrate more complex features as inter-periodic recall. This is ongoing work.

Acknowledgements This paper has benefited from a support of the Program "Complex systems in Social Sciences" (ELICCIR project) of the French National Center for Scientific Research (CNRS). We would like to thank the participants of the 10th WEHIA and of the International Conference on Cognitive Economics in Sofia for helpful comments. Ed Hopkins and an anonymous referee are gratefully acknowledged for useful suggestions and also the guest Editor, Sheri Markose, for a number of corrections. All errors remain our own.

Appendix: definition of statistics

For each simulation, we have chosen T so as to ensure that the process has converged. The three following statistics were then used:

- Posted price distribution: relative frequencies of all prices posted during the last 100 periods.
- Average posted price $\overline{p} = \frac{1}{100}\sum_{t=T-100}^{t=T}\left(\frac{1}{S}\sum_{s=1}^{S} p_{t,s}\right)$ where $p_{t,s}$ is the price posted by seller s at period t.
- Std posted price $= \frac{1}{100}\sum_{t=T-100}^{t=T}\left(\frac{1}{S}\sum_{s=1}^{S}\left(p_{t,s}-\overline{p_t}\right)^2\right)$ with $\overline{p_t} = \frac{1}{S}\sum_{s=1}^{S}\left(p_{t,s}\right)$ (average price posted at period t).
- Average price accepted by informed buyers $= \sum_{t=T-100}^{t=T}\sum_{b=1}^{I} p_{t,b}^{\text{accepted}}$ where I is the total number of informed buyers($I = aB$) and where $p_{t,b}^{\text{accepted}}$ is the price accepted by buyer b at period t.
- For the calculation of the figures used in different tables and charts, we took the average of the preceding statistics over the 25 sessions performed for all configurations.[22]

[22] For the reference configuration, Table 2 and Figs. 6 and 7 were calculated over 100 sessions whereas for the other tables we took a sample of 25 sessions for the reference configuration.

References

Beggs AW (2005) On the convergence of reinforcement learning. J Econ Theory 122:1–36
Benaïm M, Hofbauer J, Hopkins E (2005) Learning in games with unstable equilibria. Unpublished paper
Brenner T (2002) A behavioural learning approach to the dynamics of prices. Comput Econ 19:67–94
Camerer C, HO T-H (1999) Experience-weighted attraction learning in normal form games. Econometrica 67:4, 827–874
Capra MR, Goeree JK, Gomez R, Holt J (2002) Learning and noisy equilibrium behavior in a experimental study of imperfect price competition. Int Econ Rev 43(3):613–636
Darmon E, Waldeck R (2005) Convergence of reinforcement learning to Nash equilibrium: a search-market experiment. Physica A 355:1
De Palma A, Thisse J-F (1987) Les Modèles de Choix Discrets. Annales d'Economie et de Statistique, 9:151–190
Erev I, Roth AE (1998) Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. Am Econ Rev 88:4, 848–881
Hopkins E, Seymour R (2002) The stability of price dispersion under seller and consumer learning. Int Econ Rev 19:1, 46–76
Kirman AP, Vriend NJ (2001) Evolving market structure: an ACE model of price dispersion and loyalty. J Econ Dyn Control 25:3–4, 459–502
Morgan J, Orzen H, Sefton M (2006) An experimental study of price dispersion. Games Econ Behav 54(1):134–158
Posch M (1997) Cycling in a stochastic learning algorithm for normal form games. J Evol Econ 7:2, 193–207
Salmon TC (2001) An evaluation of econometric models of adaptive learning. Econometrica 69:1597–1628
Simon H (1976) From substantive to procedural rationality. In: Latsis SJ Methods and appraisal in economics. Cambridge University Press, Cambridge, pp 129–148
Stigler GJ (1961) The economics of Information. J Polit Econ 69(6):213–225
Sutton RS, Barto AG (1998) Reinforcement learning, an introduction. MIT Press, Cambridge
Varian H (1980) A model of sales. Am Econ Rev 70:651–659
Waldeck R (2006) Search and price competition. J Econ Behav Organ (forthcoming)