Artificial intelligence, algorithmic pricing and collusion Online Appendix[*]

Emilio Calvano[†] Giacomo Calzolari[‡] Vincenzo Denicolò[§] Sergio Pastorello[¶]

May 2020

This document contains additional material for sections II-VI. Numeration of sections follow that of the paper, e.g. results about section III are here in section A3.

[*] Emilio Calvano gratefully acknowledges research funding from the Italian Ministry of Education, University and Research, (PRIN project 2017TMFPSH).

[†] Bologna University, Toulouse School of Economics and CEPR

[‡] European University Institute, Bologna University and CEPR (Corresponding author)

[§] Bologna University and CEPR

[¶] Bologna University

A3 Outcomes

A3.1 Convergence

Figure A1 shows the number of iterations required to achieve convergence, as defined in the text. Convergence typically takes a large number of iterations, being achieved practically only when exploration has already faded away almost completely. The number of iterations ranges from 400,000 when β is largest (and hence exploration is limited) to several millions when exploration is very extensive.

A3.2 Profits

Figure A2 shows the prices observed upon convergence. Typically, prices are somewhat below the monopoly level but substantially above the one-shot Bertrand equilibrium.

Figure A3 shows the fraction of sessions in which the algorithms settle to a constant price (not necessarily the same for both algorithms). The other sessions converge to cycles, most of which have a period equal to 2.

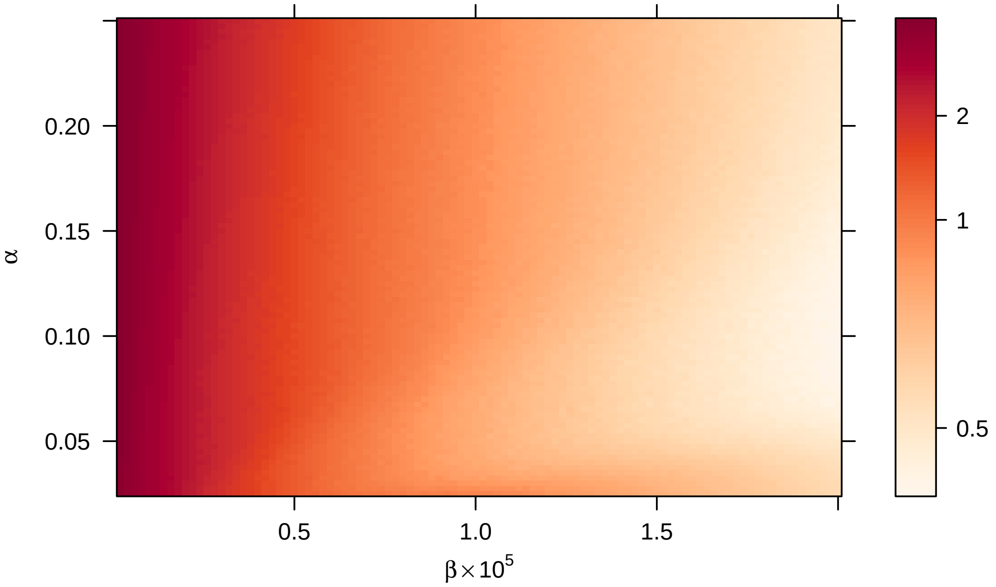


Figure A1: Number of iterations (in millions) needed to achieve convergence for a grid of values of α and β.

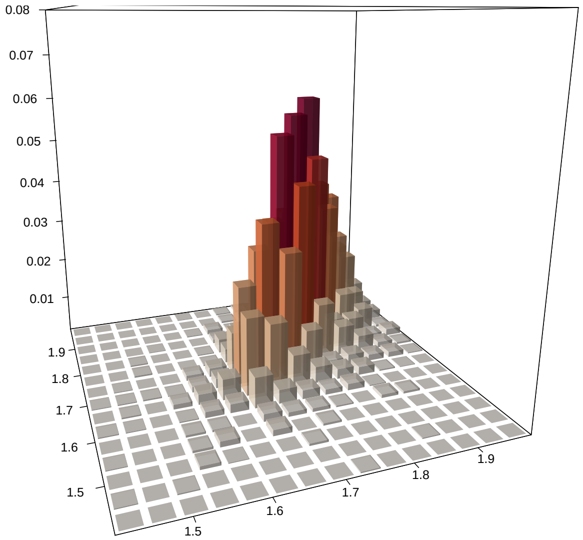


Figure A2: Long-run prices in our representative experiment. The one-shot Bertrand equilibrium price lies somewhere in between the second and third lowest prices, the monopoly price between the second and third highest.

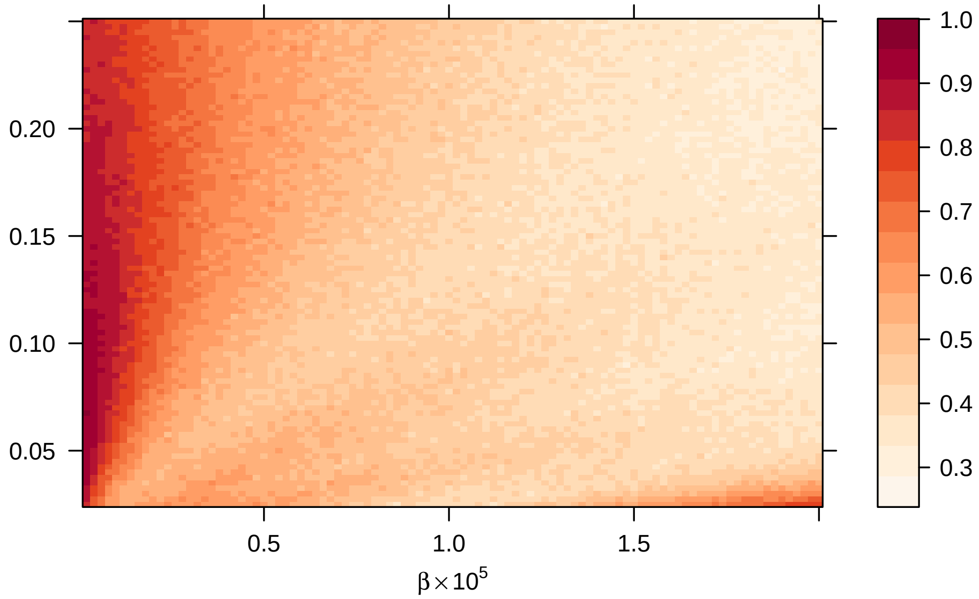


Figure A3: Fraction of sessions in which the algorithms settle to a constant price (not necessarily the same for both algorithms).

A3.3 Equilibrium play

Figure A4 and A5 compare the algorithms' limit strategies to the true best response to the rival's limit strategy. In particular, Figure A4 shows the Q-loss from not playing a best response on path, i.e. in those states that are actually reached upon convergence. This is a measure of how far the algorithms are from playing a Nash equilibrium. It is equal to zero when the algorithms are best responding, strictly positive otherwise. Figure A5 provides the same information for all states. Thus, it provides a measure of the distance from a sub-game perfect equilibrium. In fact, the algorithms almost never learn to play a sub-game perfect equilibrium. However, one can count the number of states in which they are playing a best response. This information is represented in Figure A6, showing the fraction of states for which at least one algorithm is playing a best response, and A7, showing the fraction of states where both algorithms are best responding to the rival.

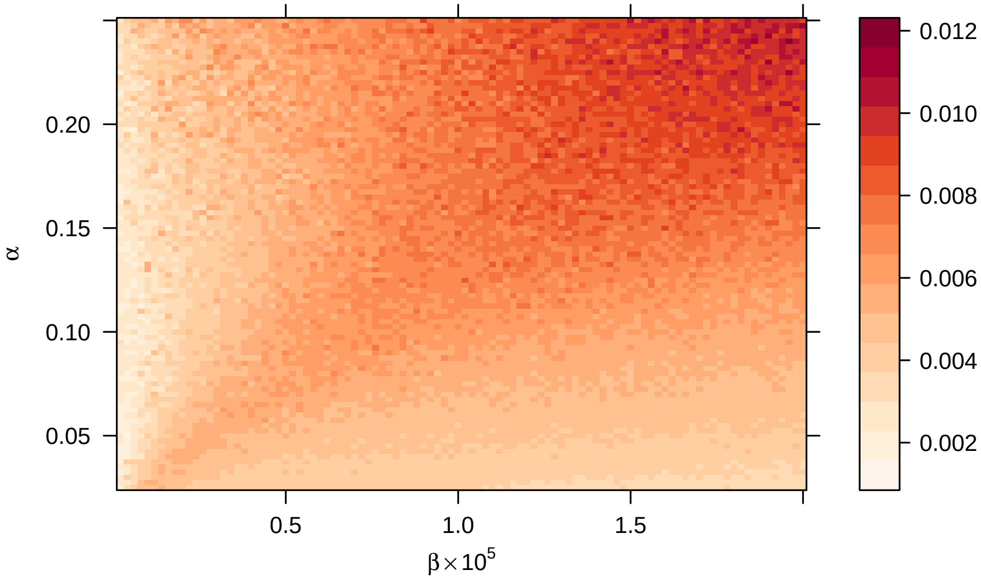


Figure A4: The average Q-loss for states that are reached on path, for a grid of values of α and β. The Q-Loss is the percentage difference between the maximum theoretical payoff that could be achieved by playing a best response to the rival's limit strategy and the actual payoff achieved by playing the limit strategy.

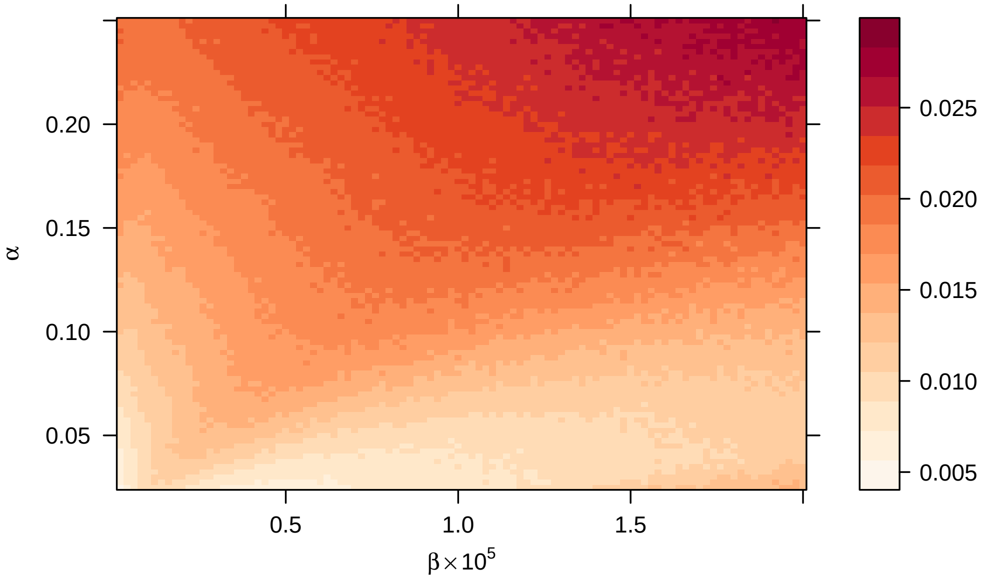


Figure A5: The average Q-Loss for all states, for a grid of values of α and β.

A4 Anatomy of collusion

The results in this section refer to the representative experiment with $\alpha = 0.15$ and $\beta = 4 \times 10^{-6}$ discussed in section 5, unless differently specified.

A4.1 Competitive environments

With memoryless algorithms ($k = 0$), there is no loss of generality in setting the discount factor δ to 0. Furthermore, the Q-matrix is much smaller than in the baseline case, as it reduces to a vector of 15 elements. This means that we can greatly speed up the learning process, which now requires much less exploration. For example, to get the same value of ν as in our representative experiment, one must set $\beta = 10^{-4}$. And we have already noted that with less exploration, the learning parameter α can be safely raised, which further increases the speed of learning. Thus, Table A1 reports the results for the case $k = 0$ with $\delta = 0$, $\beta = 10^{-4}$ and $\alpha = 0.25$: for these values, convergence is achieved in just 5,000 periods.

In this relatively short time span, the algorithms do not learn perfectly to play the unique sub-game perfect Nash equilibrium, which is to play the one-shot Nash equilibrium in all periods, but come quite close. The Nash equilibrium is played in slightly more than a quarter of the sessions. In the other sessions, one or both algorithms would have profitable deviations, but the Q-loss is less than

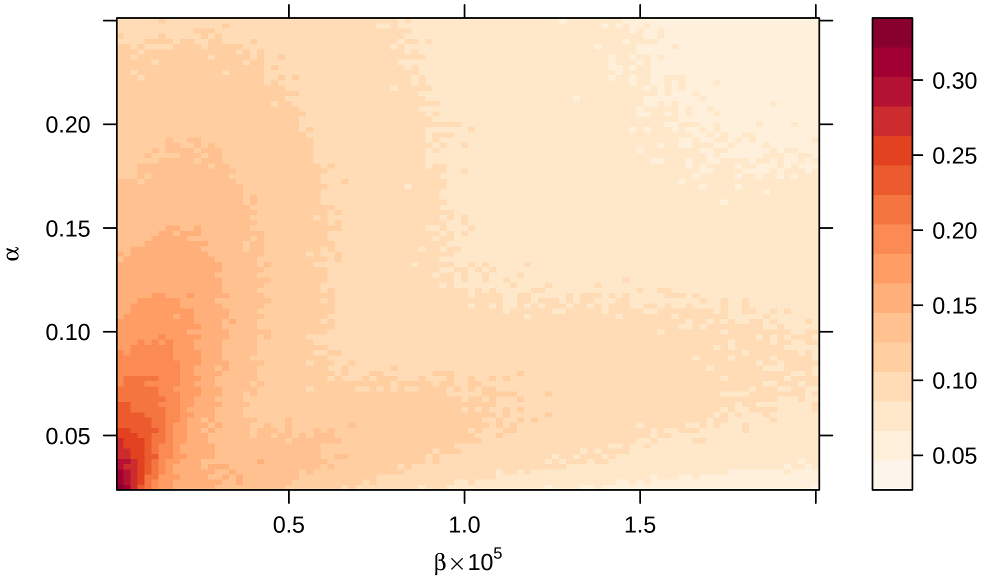


Figure A6: Fraction of all states where agents are mutually best-responding, for a grid of values of α and β.

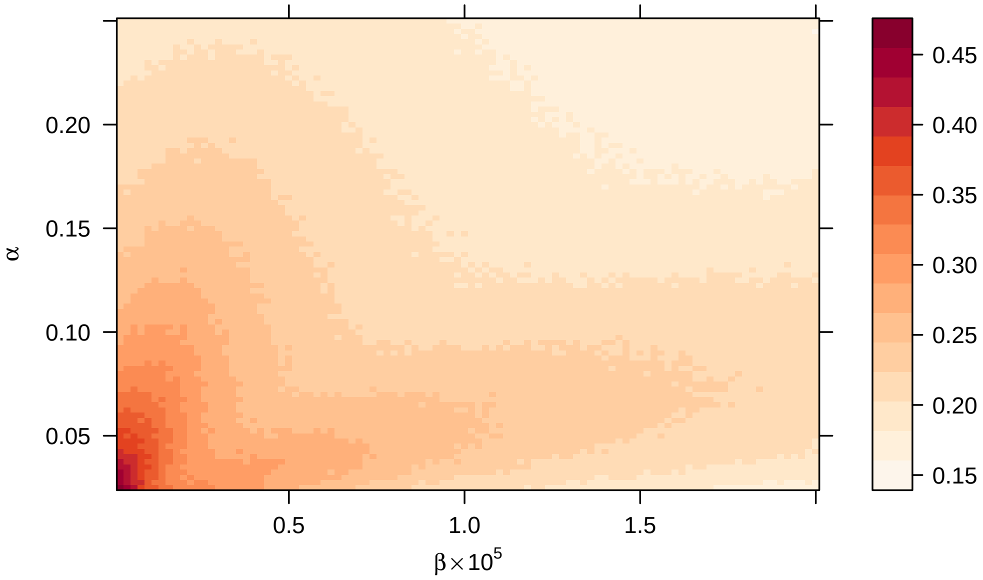


Figure A7: Fraction of all states where at least one agent's limit strategy is a best-response to the rival's one, for a grid of values of α and β.

Table A1:

Average Profit Gain	Average Equilibrium Play	Average Q-Loss
0.185	0.273	0.013

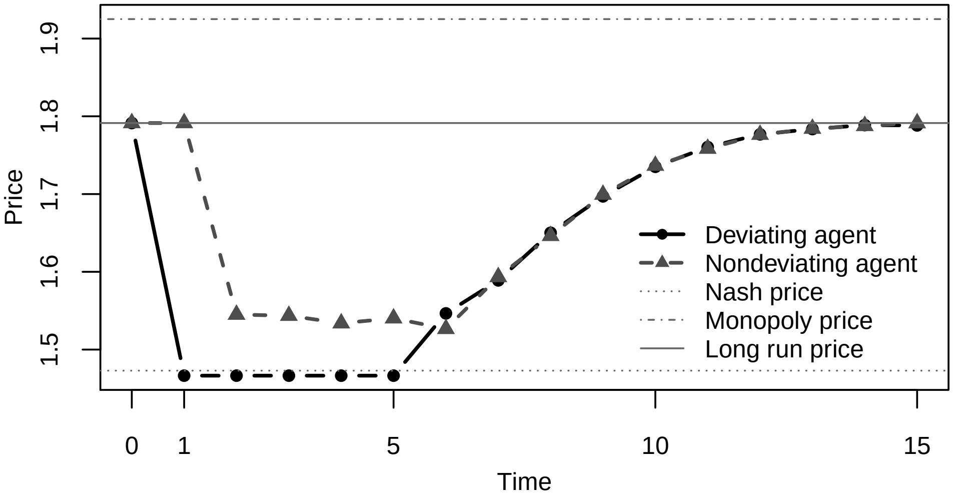


Figure A8: Impulse-response functions for the case of a 5-period deviation to the one-shot Bertrand-Nash equilibrium price (average across all sessions of our representative experiment).

2%. Overall, the average profit gain is less than 20%, i.e. about 5% above what it would be if the Bertrand equilibrium price in continuous action space were simply approximated by excess rather than by defect.

A4.2 Deviations and punishments

Figure A8 shows the average impulse response function for 5-period deviations to the one-shot Bertrand-Nash equilibrium price. The qualitative shape is similar to that observed in the case of one-period deviations.

Going back to the case of one-period deviations discussed in the main text, Figure A9 shows more moments of the distribution of the impulse responses. As explained in footnote 33, attention is restricted to sessions that converge to a constant pair of prices rather than a cycle. In spite of the heterogeneity across sessions, Figure A9 confirms that the pattern of punishment is robust.

Table A2 reports the prices charged by the two algorithms immediately after the defection (i.e., in period $\tau = 2$).

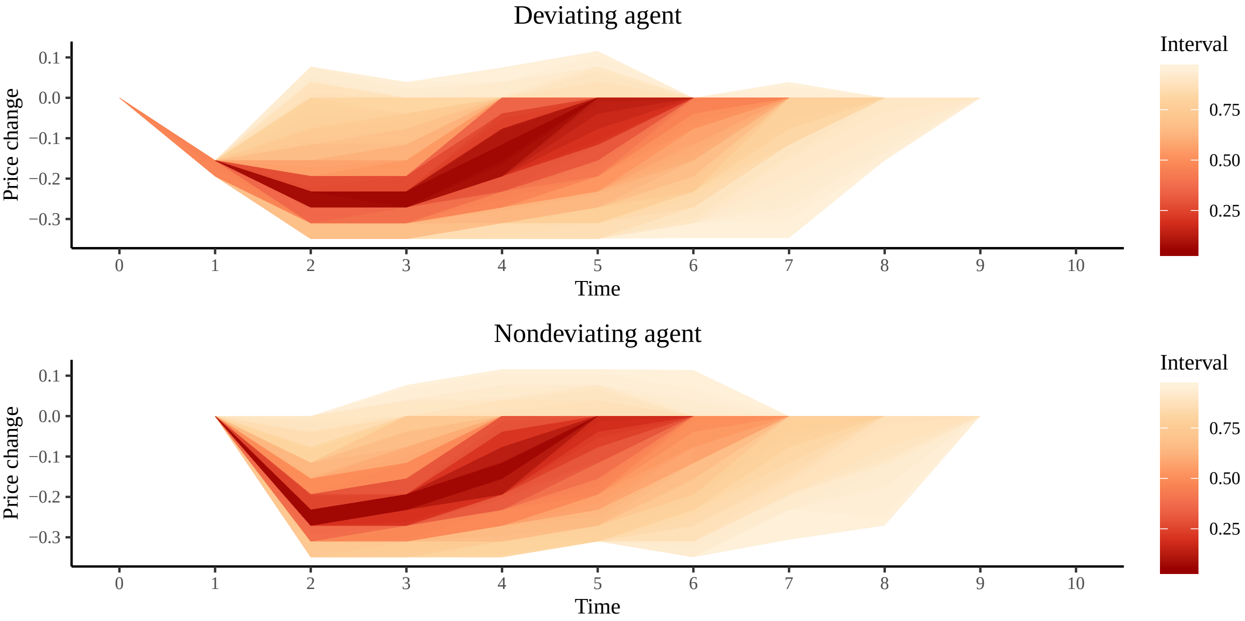
Table A2: Price changes after deviation

Panel a: Relative price change by the non deviating agent in period $\tau = 2$

Pre-shock		Deviation price														
price	Freq.	1.43	1.47	1.51	1.54	1.58	1.62	1.66	1.70	1.74	1.78	1.82	1.85	1.89	1.93	1.97
1.62	0.01	-0.04	-0.08	-0.07	-0.04	-0.08	0									
1.66	0.06	-0.08	-0.09	-0.09	-0.09	-0.08	-0.08	0								
1.70	0.11	-0.10	-0.09	-0.10	-0.10	-0.10	-0.10	-0.10	0							
1.74	0.16	-0.11	-0.11	-0.12	-0.11	-0.12	-0.11	-0.11	-0.11	0						
1.78	0.19	-0.13	-0.13	-0.13	-0.13	-0.13	-0.13	-0.13	-0.12	-0.13	0					
1.82	0.18	-0.15	-0.15	-0.14	-0.15	-0.14	-0.14	-0.14	-0.14	-0.14	-0.14	0				
1.85	0.11	-0.16	-0.16	-0.17	-0.17	-0.16	-0.16	-0.15	-0.15	-0.15	-0.16	-0.15	0			
1.89	0.09	-0.18	-0.18	-0.17	-0.18	-0.16	-0.17	-0.16	-0.16	-0.17	-0.16	-0.17	-0.16	0		
1.93	0.05	-0.19	-0.20	-0.19	-0.17	-0.19	-0.17	-0.18	-0.17	-0.18	-0.18	-0.18	-0.18	-0.16	0	
1.97	0.03	-0.19	-0.20	-0.21	-0.21	-0.21	-0.21	-0.18	-0.17	-0.17	-0.19	-0.18	-0.18	-0.17	-0.18	0

Panel b: Relative price change by the deviating agent in period $\tau = 2$ with respect to $\tau = 1$

Pre-shock		Deviation price														
price	Freq.	1.43	1.47	1.51	1.54	1.58	1.62	1.66	1.70	1.74	1.78	1.82	1.85	1.89	1.93	1.97
1.62	0.01	0.06	0.04	0.04	-0.01	-0.05	0									
1.66	0.06	0.07	0.06	0.01	-0.01	-0.02	-0.05	0								
1.70	0.11	0.08	0.06	0.04	0.01	-0.02	-0.03	-0.07	0							
1.74	0.16	0.09	0.07	0.03	0.02	-0.01	-0.04	-0.05	-0.08	0						
1.78	0.19	0.09	0.06	0.03	0	-0.02	-0.04	-0.05	-0.08	-0.11	0					
1.82	0.18	0.09	0.07	0.04	0.01	0	-0.03	-0.05	-0.07	-0.09	-0.12	0				
1.85	0.11	0.09	0.08	0.03	0.01	-0.01	-0.02	-0.04	-0.07	-0.09	-0.11	-0.12	0			
1.89	0.09	0.10	0.08	0.03	0.01	0	-0.03	-0.04	-0.06	-0.08	-0.11	-0.12	-0.14	0		
1.93	0.05	0.10	0.07	0.05	0.01	0.01	-0.02	-0.04	-0.07	-0.09	-0.09	-0.11	-0.15	-0.16	0	
1.97	0.03	0.13	0.10	0.07	0.02	0	-0.03	-0.02	-0.04	-0.06	-0.10	-0.11	-0.12	-0.13	-0.18	0


Deviating agent

Nondeviating agent

Figure A9: Fan chart of impulse responses, for sessions converging to a constant pair of prices. The variable on the vertical axis is the percentage price change relative to the long-run prices.

Table A3 shows that deviations considered in table A3 are almost always unprofitable.

Table A4 reports the length of the punishment phase for a range of long-run prices and price cuts. The duration of the punishment is not very sensitive to these variables.

Table A3: Unprofitability of deviations

Panel a: Average percentage gain from the deviation in terms of discounted profits

Pre-shock		Deviation price														
price	Freq.	1.43	1.47	1.51	1.54	1.58	1.62	1.66	1.70	1.74	1.78	1.82	1.85	1.89	1.93	1.97
1.62	0.01	-0.03	-0.02	-0.02	-0.01	-0.02	0.00									
1.66	0.06	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	0.00								
1.70	0.11	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	0.00							
1.74	0.16	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	0.00						
1.78	0.19	-0.04	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	0.00					
1.82	0.18	-0.04	-0.04	-0.04	-0.03	-0.03	-0.03	-0.03	-0.03	-0.04	-0.04	0.00				
1.85	0.11	-0.04	-0.04	-0.04	-0.04	-0.03	-0.03	-0.03	-0.04	-0.03	-0.04	-0.04	0.00			
1.89	0.09	-0.04	-0.04	-0.04	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.04	-0.04	0.00		
1.93	0.05	-0.04	-0.04	-0.04	-0.04	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.04	-0.04	0.00	
1.97	0.03	-0.04	-0.04	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.04	0.00

Panel b: Frequency of unprofitable deviations.

Pre-shock		Deviation price														
price	Freq.	1.43	1.47	1.51	1.54	1.58	1.62	1.66	1.70	1.74	1.78	1.82	1.85	1.89	1.93	1.97
1.62	0.01	1.00	0.91	0.91	0.82	0.91	0.00									
1.66	0.06	1.00	0.96	0.95	0.95	0.95	0.97	0.00								
1.70	0.11	0.99	0.98	0.99	0.98	0.98	0.99	0.98	0.00							
1.74	0.16	0.99	0.99	0.97	0.97	0.95	0.95	0.95	0.97	0.00						
1.78	0.19	0.99	0.99	0.98	0.97	0.96	0.97	0.98	0.97	0.98	0.00					
1.82	0.18	0.99	1.00	0.98	1.00	0.99	0.99	0.97	0.97	0.98	0.99	0.00				
1.85	0.11	0.99	0.98	0.97	0.97	0.97	0.98	0.99	0.99	0.96	0.98	0.97	0.00			
1.89	0.09	0.98	0.98	0.97	0.97	0.95	0.94	0.97	0.94	0.96	0.95	0.97	0.98	0.00		
1.93	0.05	0.99	0.97	1.00	0.97	0.93	0.97	0.97	0.99	0.99	0.94	0.97	1.00	0.96	0	
1.97	0.03	1.00	1.00	1.00	0.97	0.97	0.94	0.88	0.94	0.91	0.91	0.94	0.97	0.94	1	0

Table A4: Average length of punishment

Pre-shock		Deviation price														
price	Freq.	1.43	1.47	1.51	1.54	1.58	1.62	1.66	1.70	1.74	1.78	1.82	1.85	1.89	1.93	1.97
1.62	0.01	5.18	4.27	4.91	4.27	5.00	1.00									
1.66	0.06	5.24	4.96	5.35	5.04	4.88	4.95	1.00								
1.70	0.11	5.47	5.42	5.55	5.49	5.54	5.55	5.63	1.00							
1.74	0.16	5.90	5.86	5.85	5.70	5.68	5.79	5.50	5.58	1.00						
1.78	0.19	5.86	5.72	5.71	5.69	5.70	5.85	5.57	5.65	5.63	1.00					
1.82	0.18	6.22	6.07	6.22	6.23	6.12	6.26	6.15	6.17	6.06	6.00	1.001				
1.85	0.11	6.67	6.51	6.40	6.62	6.55	6.34	6.27	6.51	6.30	6.36	6.24	1.00			
1.89	0.09	6.52	6.26	6.48	6.40	6.58	6.34	6.28	6.53	6.36	6.35	6.33	6.58	1.00		
1.93	0.05	6.75	6.54	7.13	6.18	6.00	6.57	6.91	6.63	6.69	6.51	6.33	6.37	6.33	1.00	
1.97	0.03	6.61	6.88	6.39	6.45	6.30	6.45	6.70	6.79	6.12	6.33	6.18	5.88	6.33	6.85	1.00

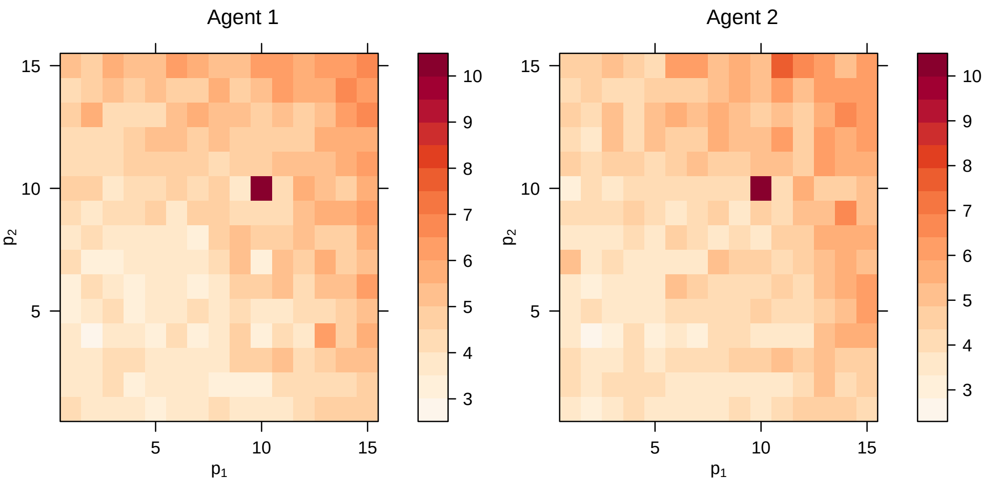


Figure A10: Average best response functions. Prices are denoted by their position in the discretized grid: for example, the 3rd price is 1.51, while 10th price is 1.78.

A4.3 The graph of strategies

Figure A10 depicts the average best response functions, for those sessions where the algorithms converge to the same pair of supra-competitive prices. In particular, the figure represents case in which both algorithms converge to the 10-th price of the grid, which is the most frequent outcome. The average function F exhibits a spike at that point. Elsewhere prices are much lower. This reflects the punishment of deviations. The punishment is harshest for large price cuts, and less harsh in case of upwards deviations. But apart from that, there emerges no recognizable pattern.

We next provide support to the claim that cooperation eventually resumes both after unilateral and multilateral deviations, Figure A11 shows the distribution of the number of states (that is, nodes in the graph) starting from which this does not happen. In more than 90% of the session, this number is zero, and in almost 98% of the sessions, it is not greater than 3.

Figure A12 reports the histogram of the number of periods it takes to go back to the absorbing state starting from any possible state. The distribution includes all 1,000 sessions of our representative experiment, for a total of 225,000 trajectories. The median is 5 periods, and rarely it takes more than 10 periods to return to the long-run prices.

The main text claims also that there are a few key nodes that act as gateways, either directly or indirectly, to the absorbing node. This claim can be supported by measuring a node's centrality and then looking at the concentration of centrality. We measure a node's centrality by its betweenness with respect to the limit path; in other words, the centrality of a node is given by the number of shortest paths that pass through it between any node in the network and one of the limit path nodes. Figure A13 illustrates the concentration of node centrality by means of the concentration-ratios curve. This curve represents the fraction of total betweenness centrality associated with a

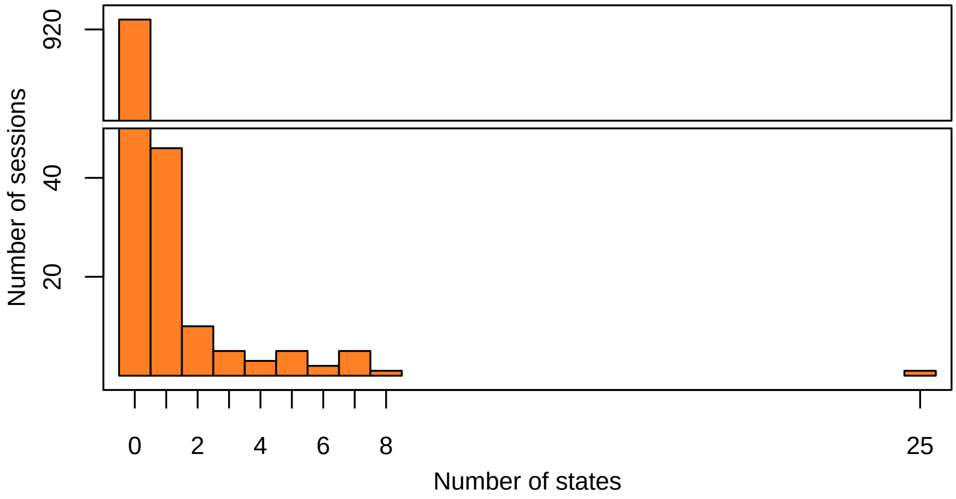


Figure A11: Distribution of the number of states from which the system does not return to the long-run prices. In our baseline experiment, there are 225 states. The distribution includes all 1,000 sessions of the experiment.

varying number of most central nodes. As shown, the five most central nodes account for more than half of total centrality.

A5 Robustness

This section reports evidence backing up the claims made in Section 6 of the text. The following tables typically report the same statistics for each experiment considered: the profit gain (denoted by Δ), the fraction of sessions converging to a Nash equilibrium, and three variables that capture the algorithms' reaction to exogenous price cuts in period $\tau = 1$ (specifically, a deviation to the static best-response to the rival's price). These variables are the average price drop by the non-deviating algorithm in period $\tau = 2$ (denoted by IR), the fraction of sessions in which the punishment makes the deviation unprofitable (denoted by IC), and the length of the punishment.

A5.1 Number of players

Table A5 reports the results of experiments with three or four firms.

As noted in the main text, in interpreting the results reported in Table A3 one must keep in mind that with $n = 3$ or $m = 4$ the Q-matrix becomes significantly larger, which means that holding β constant the effective level of experimentation is in fact much lower. Table A6 describes, for the case $n = 3$, the effect of decreasing β so as to achieve a level of exploration comparable to

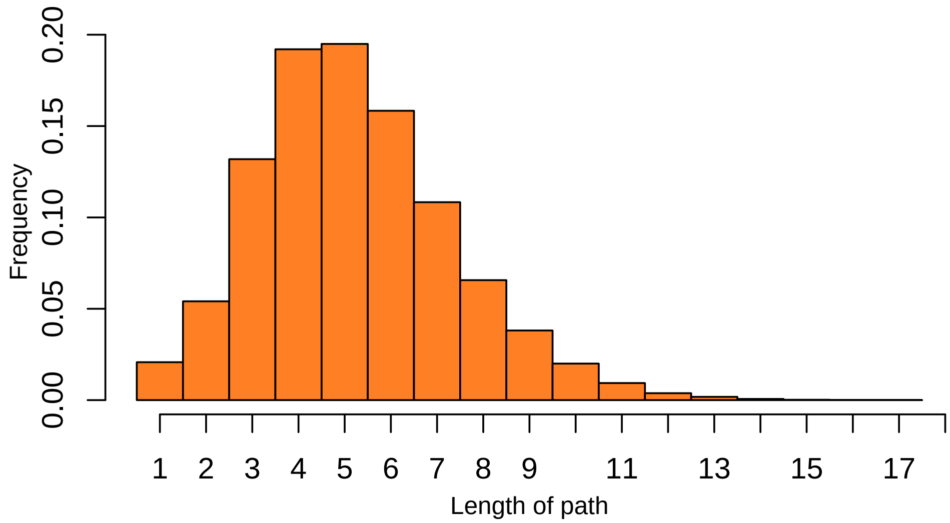


Figure A12: Distribution of the length of the path back to the long-run prices, for all 1,000 sessions of the representative experiment. The distribution includes only those nodes starting from which the system returns to the long-run prices.

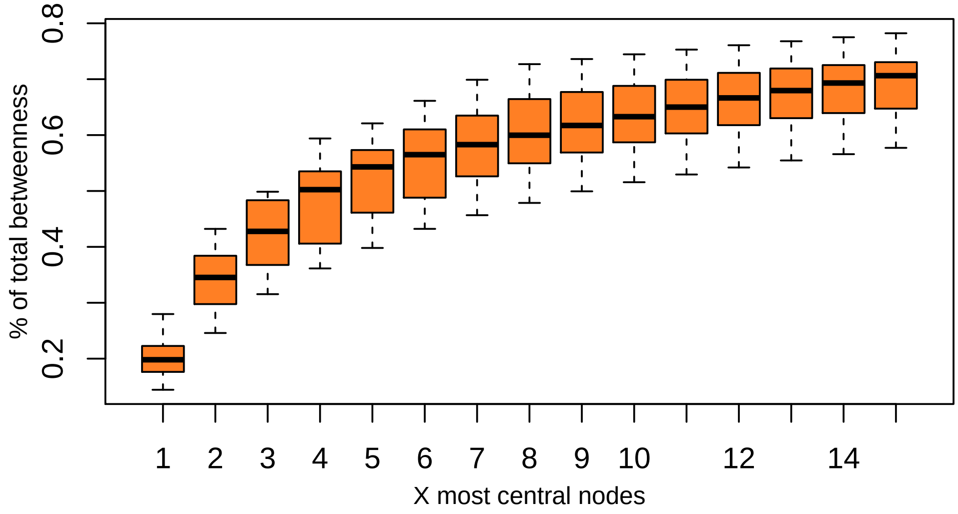


Figure A13: The concentration-ratios curve of betweenness centrality. The distribution includes only those nodes starting from which the system returns to the long-run prices. Results refer to the 1,000 sessions of the representative experiment.

Table A5

	Δ	Equil. Play on path	IR	IC	Punishment length
$n = 2$	0.849	0.505	-0.127	0.936	5.705
$n = 3$	0.643	0.184	-0.078	0.882	11.686
$n = 4$	0.559	0.717	-0.080	0.899	18.379

the case of duopoly, with a corresponding adjustment in the learning rate (with more exploration, learning must be more persistent to be equally effective, so α should be lowered). Table A4 confirms that increasing the level of exploration restores at least part of the profit gain that is lost when the number of players increases. We observe also an increase in the rate of equilibrium play, and stronger but shorter (and, in the end, more effective) punishments.

Table A6:

				Equilibrium Play			Punishment
n	α	β	Δ	on Path	IR	IC	Length
3	0.15	0.4	0.643	0.184	-0.078	0.882	11.686
3	0.05	0.024	0.750	0.306	-0.146	0.908	7.548

Notes: The values of β_2 are premultiplied by 10^5.

Figure A14 shows the average impulse-response function for the case of three firms; the case of four firms is similar. Clearly, deviations are punished. In period $\tau = 2$, the intensity of the punishment is similar to the case of duopoly, but in subsequent periods cooperation re-starts much more slowly. Overall, then, the punishment seems more intense than in a duopoly.

A5.2 Asymmetric firms

The main text reports the effects of cost asymmetries. Table A7 instead focuses on demand asymmetries. In particular, it considers the case in which the product supplied by firm 2 is of greater quality, and hence in higher demand, than that of firm 1. In addition to the standard variables, the table reports also an index of asymmetry (namely, firm 2's market share in the Bertrand equilibrium of the one-shot game) and an index of the way the gains from cooperation are divided between the two firms. The table confirms that cost and demand asymmetries have almost identical effects.

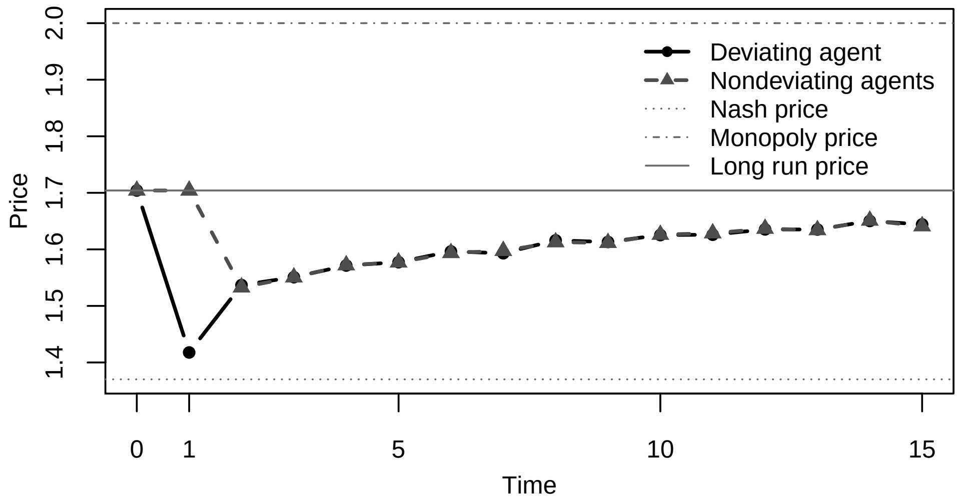


Figure A14: Average impulse-response function for the case of three firms

Table A7

a_2	Agent 2's Nash market share	Δ	$\frac{\pi_1/\pi_1^N}{\pi_2/\pi_2^N}$	Equilibrium play on path	IR	IC
2.00	0.50	0.85	1.00	0.50	-0.13	0.94
2.12	0.55	0.84	1.05	0.54	-0.15	0.92
2.25	0.59	0.81	1.12	0.52	-0.16	0.92
2.38	0.63	0.78	1.19	0.55	-0.17	0.91
2.50	0.66	0.76	1.27	0.56	-0.19	0.93
2.75	0.72	0.71	1.44	0.57	-0.21	0.89

A5.3 Stochastic demand

Table A8 reports the results for the case in which the aggregate demand parameter a_0 takes on three values, i.e. $a_0^L = -a_0^H$, 0 and a_0^H, with the same probability, thus generating both negative and positive demand shocks. The shocks are purely idiosyncratic and have no persistency. Demand variability does hinder collusion (lower profit gains), as one would have expected, but it does not eliminate it. We observe less equilibrium play on path. As for the punishments, intensity and duration are almost unaffected but overall the punishment seems to be less effective.

Table A8:

a_0^H	Δ	Equilibrium Play on Path	IR	IC	Punishment Length
0.00	0.848	0.509	−0.125	0.916	5.708
0.05	0.848	0.485	−0.129	0.906	5.657
0.10	0.832	0.462	−0.134	0.896	5.792
0.15	0.797	0.433	−0.134	0.879	5.737
0.20	0.750	0.407	−0.130	0.867	5.726
0.25	0.695	0.337	−0.127	0.842	5.675

A5.4 Variable market structure

Table A9 reports the results for the case of serially correlated entry or exit. The parameter ρ is the probability of entry when there are two active firms, and of exit when there are three active firms.

The average profit gain decreases to 58% and 56% when ρ is set at 0.1% or at 0.01%, respectively. Equilibrium play is observed on path in less than 5% of the sessions in the first case, and about 7% of the sessions in the second. In both cases, punishments are substantially milder than in the baseline experiment, and this makes exogenous deviations profitable in more than 50% of the cases.

Table A9

ρ	Δ	Equilibrium Play on Path	IR	IC	Punishment Length
0.001	0.583	0.048	−0.006	0.432	9.620
0.0001	0.566	0.070	−0.009	0.450	9.829

A5.5 Product substitutability

Table A10 explores the effects of changing the parameter μ, which is an index of the degree of product differentiation. In the baseline model, we have $\mu = 0.25$. Here we let the parameter range from $\mu = 0.5$ to $\mu = 0$ (the case of perfect substitutes). For this latter case, the system of demand functions is obtained by taking the limit for $\mu \to 0$ and using L'Hopital's rule.

The profit gain slightly decreases with product substitutability but still remains quite high even in the case of perfect substitutes. The punishment price drops significantly and the punishment length slightly decreases. Equilibrium play is highest in the case of perfect substitutes.

Table A10

μ	Δ	Equilibrium Play on Path	IR	IC	Punishment Length
0.50	0.856	0.331	−0.095	0.906	6.170
0.45	0.860	0.371	−0.099	0.912	6.063
0.40	0.859	0.405	−0.102	0.914	5.954
0.35	0.853	0.414	−0.108	0.912	5.944
0.30	0.859	0.433	−0.118	0.919	5.804
0.25	0.849	0.505	−0.127	0.936	5.705
0.20	0.839	0.532	−0.139	0.918	5.562
0.15	0.835	0.582	−0.157	0.911	5.449
0.10	0.829	0.660	−0.171	0.897	5.465
0.05	0.827	0.694	−0.201	0.919	5.677
0.01	0.805	0.654	−0.220	0.875	5.560
0.00	0.774	0.631	−0.214	0.879	5.484

A5.6 Alternative initializations

We consider different specifications of the $\mathbf{Q}_0$ matrix. Remember that in our baseline experiments we initialize the Q-matrix on the assumption that the rival randomizes uniformly across the m feasible prices, which implies

$$Q_{i,0}(s,a_i) = \frac{\sum\limits_{a_{-i}\in A^{n-1}} \pi_i(a_i,a_{-i})}{(1-\delta)\,|A|^{n-1}}$$

One alternative is to assume instead that the rival always charges a constant price $\bar{p}$. In this case, we simply set

$$Q_{i,0}(s,a_i) = \frac{\pi_i(a_i,\bar{p})}{1-\delta}.$$

In particular, we have implemented this initialization setting $\bar{p}$ at the static equilibrium Bertrand price. This is best approximated (by excess) by the third lowest price of our grid. The alternative of using the second lowest price (approximation by defect) leads to very similar results.

We have also considered the case where the rival is playing a grim-trigger strategy (that is $p_{-i,t} = p^H$ if $p_{i,t-1} = p_{-i,t-1} = p^H$ and $p_{-i,t} = p^L$ otherwise, with some prices $p^H > p^L$). When δ is large, the corresponding initial matrix may be approximated by

$$Q_{i,0}(s,a_i) = \begin{cases} \frac{\pi_i(a_i,p^H)}{1-\delta} \text{ if } a_i = p^H \\ \frac{\pi_i(a_i,p^L)}{1-\delta} \text{ if } a_i \neq p^H. \end{cases}$$

(In fact, our simulations use the exact formulas that accounts also for the profits obtained in the first period.) We have implemented this initialization setting p^H at the monopoly level and p^L at the static equilibrium Bertrand level. Again, these prices may be approximated by excess or by defect. The table reports the results for the closest approximation (where the Bertrand price is approximated by excess and the monopoly price by defect), but we have considered all possible combinations of approximations, obtaining very similar results.

Finally, we have considered the case in which the initial Q-matrix is constant, i.e.

$$Q_{i,0}(s, a_i) = \bar{Q}$$

with $\bar{Q} = 5$ (which is close to the average Q value in our benchmark case) and $\bar{Q} = 10$, and the case where $Q_{i,0}(s, a_i)$ is a random draw from a uniform distribution on $[0, 10]$.

Table A11

Q initialization	Δ	Equilibrium Play on Path	IR	IC	Punishment Length
Benchmark	0.849	0.505	-0.127	0.936	5.705
Nash	0.731	0.647	-0.104	0.915	5.244
Grim Trigger	0.724	0.640	-0.104	0.925	5.350
Random (0,10)	0.877	0.238	-0.113	0.853	5.604
Uniform at Q=5	0.793	0.621	-0.119	0.928	5.481
Uniform at Q=10, no expl.	0.978	0.525	-0.120	0.951	9.619

Our results are rather robust with respect to these changes in the initialization of the Q-matrix. The profit gain is lowest when the Q-matrix is initialized at Nash, or at grim-trigger strategies. However, in both cases Δ exceeds 70% and the impulse response functions are very similar to the baseline case. By way of contrast, collusion is almost perfect when when $\mathbf{Q}_0$ is a constant matrix and $\bar{Q} = 10$. In this case, even if exploration is shut down, the algorithms visit systematically all cells of the matrix and thus are able to learn in a very effective way.

A5.7 Alternative action sets

Table A12 shows the results obtained by varying the grid of feasible prices. First, we have considered an enlargement of the grid while keeping the number of feasible prices constant at $m = 15$. We have considered both a symmetric enlargement, which simply amounts to an increase in ξ, and one biased downwards. In this latter case, the lowest price is set at $p = 0.99$, i.e. just below the marginal cost. Next, we have increased the number of feasible prices from $m = 15$ to $m = 50$ and $m = 100$.

When m is kept constant, increasing ξ has a limited impact. The effect of using a price grid biased downwards is more substantial, but the profit gain is still close to 60%. With a larger action space

Table A12:

Number of prices	Δ	Equilibrium Play on Path	IR	IC	Punishment Length
15 (benchmark)	0.849	0.505	-0.127	0.936	5.705
15, ξ=0.5	0.761	0.730	-0.125	0.947	4.694
15, lowest price set at 0.99	0.612	0.871	-0.112	0.942	4.258
50, ξ=0.1	0.709	0.154	-0.059	0.931	12.713
100, ξ=0.1	0.704	0.684	-0.071	0.986	23.301

the average profit gain is still around 70%. In these cases, however, the Q-matrix is substantially larger than in the baseline case, so learning off path is inevitably more limited.

A5.8 Memory

Table A13 reports the results of our simulations for the case of two-period memory ($k = 2$).

Table A13

k	Δ	Equilibrium Play on Path	IR	IC	Punishment Length
1	0.849	0.505	-0.127	0.936	5.705
2	0.574	0.371	-0.030	0.441	24.038

With $n = 2$ and $k = 2$, the Q-matrix is as large as with one-period memory and four active firms. Like in that case, some of the profit gain can be restored by allowing for more experimentation.

A5.9 Linear demand

We repeated the analysis for the case of duopoly with linear demand functions derived from a quadratic utility function of the Singh and Vives (1984) type, i.e.

$$u = q_1 + q_2 - \frac{1}{2}(q_1^2 + q_2^2) - \gamma q_1 q_2$$

for various values of the horizontal differentiation parameter γ. The results are reported in Table A14. The average profit gain is non-monotone in γ: it is well above 80% when the products are good substitutes or fairly independent, but reaches a minimum for intermediate degrees of product differentiation (to be precise, the average profit gain is 63% when $\gamma = \frac{3}{4}$).

Table A14

		Equilibrium Play			Punishment
γ	Δ	on Path	IR	IC	Length
0.01	0.856	0.224	−0.003	0.666	6.445
0.05	0.843	0.215	−0.013	0.874	6.305
0.10	0.843	0.235	−0.027	0.865	6.357
0.15	0.839	0.219	−0.042	0.867	6.367
0.20	0.837	0.249	−0.058	0.884	6.391
0.25	0.823	0.265	−0.074	0.877	6.270
0.30	0.816	0.274	−0.090	0.873	6.210
0.35	0.807	0.270	−0.109	0.871	6.208
0.40	0.808	0.282	−0.131	0.886	6.274
0.45	0.791	0.316	−0.154	0.884	6.284
0.50	0.770	0.318	−0.173	0.885	6.305
0.55	0.742	0.315	−0.188	0.865	5.988
0.60	0.725	0.343	−0.210	0.866	5.985
0.65	0.686	0.356	−0.232	0.859	5.892
0.70	0.651	0.406	−0.255	0.848	5.650
0.75	0.634	0.429	−0.289	0.849	5.723
0.80	0.649	0.417	−0.329	0.809	5.724
0.85	0.677	0.396	−0.341	0.767	5.584
0.90	0.753	0.385	−0.345	0.715	5.256
0.95	0.854	0.412	−0.410	0.804	5.603
0.99	0.871	0.389	−0.380	0.783	5.687

A5.10 Boltzmann experimentation

We have repeated our experiments for algorithms that explore according to the Boltzmann model (see footnote 13). For consistency with the baseline analysis, in which exploration diminishes as time passes, we have let the algorithms' "temperature" decrease over time according to

$$T_t = \lambda_0 \times t^{\lambda_1}$$

where we used $\lambda_0 = 1000$ everywhere. Like we did for the ε-greedy model, we have considered a grid of possible values of α and λ_1. Specifically, for α we have considered the same range as in our baseline analysis, whereas λ_1 varies from 0.999 to 0.999999596. Figure A15 shows the level of the average profit gain obtained, upon convergence, in our grid. The level of collusion compares to that of the baseline case. Like in the ε-greedy model, more exploration facilitates collusion, but now it seems that more persistent learning (i.e., lower values of α) has an unambiguous negative effect. This is probably due to the fact that exploration is no longer purely random. This guarantees that the learning process is sufficiently persistent even for larger values of α.

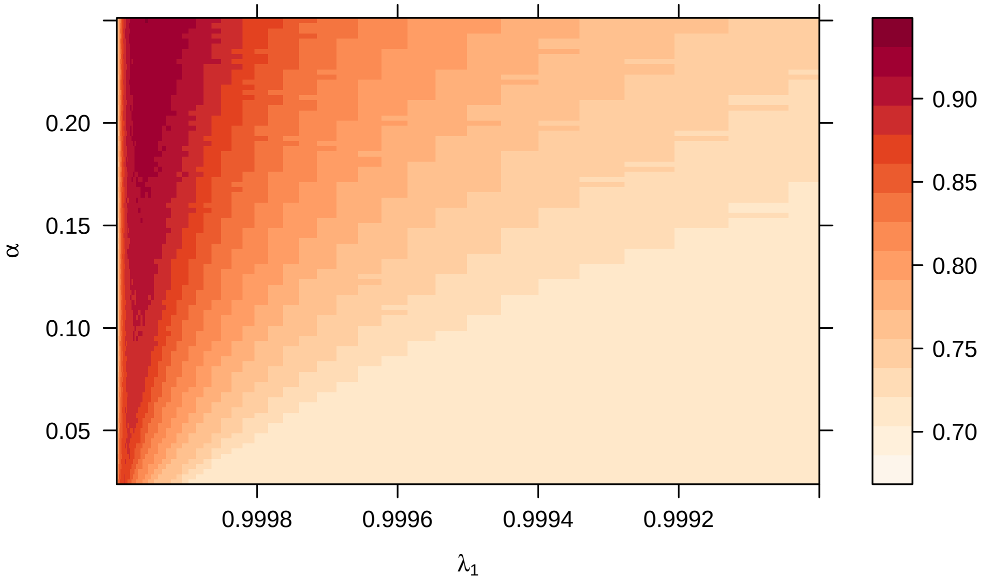


Figure A15: The average profit gain for a range of values of α and λ_1 with Boltzmann exploration.

Table A15 singles out one experiment that leads to a level of the profit gain similar to that of our representative experiment for the ε-greedy model. There are several notable differences. First, convergence is much faster, being achieved on average in less than 400,000 repetitions (this contrasts with the more than 1,500,000 repetitions required on average to achieve convergence in our representative baseline experiment). Second, there is much less equilibrium play (but again, the loss from not playing a true best response is less than 1%). Third, the punishment of deviations is less harsh, implying that now unilateral price cuts can be profitable in more than 15% of the sessions. Overall, these results suggest that with Boltzmann exploration learning is faster but less complete than in the ε-greedy model.

Table A15

α	λ_0	λ_1	Δ	Equilibrium Play on Path	IR	IC	Punishment Length
0.05	1000	0.999961	0.856	0.094	−0.090	0.841	6.554

A5.11 Asymmetric learning

Table A16 reports the results for the case in which the two algorithms have different learning rates α, or different levels of experimentation β. In all cases, the α and β parameters are kept at their baseline values for algorithm 1 and varied for algorithm 2 (the first line of the table is benchmark).

Collusion appears to be robust to these changes. The average profit gain is lower when the algorithms are asymmetric, but the effect is modest. Interestingly, the algorithm that updates more slowly α gains more, whereas the algorithm that explores more under-performs relative to the other.

Table A16

			Equilibrium Play			Punishment
α_2	β_2	Δ	on Path	IR	IC	Length
0.15	0.40	0.849	0.505	−0.127	0.936	5.705
0.05	0.40	0.821	0.304	−0.114	0.864	5.206
0.30	0.40	0.797	0.523	−0.112	0.900	6.060
0.15	0.20	0.771	0.423	−0.112	0.882	5.333
0.15	0.80	0.798	0.310	−0.109	0.853	5.599

Notes: The values of β_2 are premultiplied by 10^5.

A6 The time scale of learning

Table A17 confirms that if exploration is reactivated after the algorithms have been re-matched as described in the main text, comparable levels of collusion are achieved.

Table A17:

		Equilibrium Play			Punishment
Q initialization	Δ	on Path	IR	IC	Length
Benchmark	0.849	0.505	−0.127	0.936	5.705
Rematching	0.837	0.541	−0.128	0.932	5.674