

# Cournot versus Walras in Dynamic Oligopolies with Memory

Carlos Alós-Ferrer

Department of Economics, University of Vienna  
Hohenstaufengasse, 9. A-1010 Vienna (Austria)

## Abstract

This paper explores the impact of memory in Cournot oligopolies where firms learn through imitation of success as suggested in Alchian (1950) and modeled in Vega-Redondo (1997). As long as memory includes at least one period, the long-run outcome corresponds to any quantity between the Cournot one and the Walras one. The (evolutionary) stability of the walrasian outcome relies on inter-firm comparisons of simultaneously observed profits (i.e. whether a firm earns more than others in a given period), whereas the stability of the Cournot-Nash equilibrium is derived from intertemporal comparisons of profits for each given firm (i.e. whether a deviation pays off for that firm).

*Journal of Economic Literature* Classification Numbers: C72, D83, L13.

**Keywords:** memory, imitation, mutation, Cournot, Walras.

## 1 Introduction

The learning literature initiated by Kandori, Mailath, and Rob (1993) and Young (1993) has seen a growing number of applications to oligopoly theory recently. Starting with the analysis of a Cournot oligopoly presented in Vega-Redondo (1997) (extended in different directions by Alós-Ferrer, Ania, and Vega-Redondo (1999), Tanaka (1999) and Schenk-Hoppé (2000)),<sup>1</sup> applications have been studied for Bertrand oligopolies (Alós-Ferrer, Ania, and Schenk-Hoppé (2000) and Hehenkamp (2002)), differentiated-goods oligopolies (Tanaka (2000) and Tanaka (2001)), signaling (Nöldeke and Samuelson (1997)), and insurance markets (Ania, Tröger, and Wambach (2002)).

All these models are based on the ideas pointed out by Alchian (1950). In summary, firms adapt via imitation and trial and error. Imitation is based on observed success, where “success” is defined in terms of achieving the highest profits. Trial and error is modeled through an experimentation (also called “mutation”) parameter, which gives firms a (small) probability of trying something new.<sup>2</sup>

<sup>1</sup>See also Thijssen (2001) and Schipper (2001).

<sup>2</sup>An alternative approach is presented by Rhode and Stegeman (2001), who analyze imitative (Darwinian) dynamics in a symmetric, differentiated duopoly with linear demand functions. Experimentation, however, is both frequent and potentially large in size. In their model, firms’ strategy choices cluster around a strategy profile that is not a Nash equilibrium.

The success of this approach in oligopoly models is due both to conceptual and technical reasons. Conceptually, real oligopolies are extremely complex situations where agents (firms) are bound to use simple rules of thumb (relative to the complexity of the environment) to save decision costs. Technically, the long-run outcome of the process as the experimentation parameter goes to zero (i.e. for small perturbations) can be characterized through the techniques developed in Freidlin and Wentzell (1988), which were not available when Alchian wrote his seminal paper, but nowadays can be found in several leading books on learning (see e.g. Fudenberg and Levine (1998), Samuelson (1997), or Young (1998)).

The general idea is analogous to that underlying evolutionary models, where agents obtaining higher payoffs thrive at the expense of others. In oligopoly applications, the strategies (whether quantities, prices, or insurance contracts) leading to higher profits are promptly imitated and come to dominate the population.

In an oligopoly, this idea has interesting implications. Among them is the possible selection of non-Nash equilibria, first illustrated by Schaffer (1989) and Rhode and Stegeman (2001).<sup>3</sup> In the context of  $N$ -firm Cournot oligopolies, Vega-Redondo (1997) and Alós-Ferrer, Ania, and Vega-Redondo (1999) show that the Cournot-Nash equilibrium can be quickly discarded in favor of the “walrasian” one (where firms set price equal to marginal cost) due to the effects of *spite*, which imply that, with imitation of highest profits, only *relative* payoffs matter. On the one hand, even if firms deviating from the walrasian equilibrium earn higher profits than before, those not deviating are left with even higher profits, so deviating results in a bad relative position. On the other hand, firms deviating from the Cournot-Nash equilibrium to produce the walrasian quantity achieve lower profits than before, but those not deviating are left with even lower profits. Hence, deviating leads to a good relative position.

As pointed out in Alós-Ferrer (2000), the basic evolutionary idea is not objectionable in a biological framework, where agents live one period and are replaced by their offspring. The argument is less tenable, though, when applied to learning models in economics, in particular to firms. Since firms do not die, but presumably try to learn, the effects of spite rely on the fact that previous outcomes (e.g. the profits of the Cournot-Nash equilibrium) are immediately forgotten and only the current comparison of profits matters. Besides, if the results of even the most recent period are forgotten, it is difficult to interpret the experimentation process as trial and error, since an error might only be perceived as such when compared with previous results. Hence, it is worth investigating how evolutionary models of oligopoly are affected if firms use the information gained in recent periods of play.

Alós-Ferrer (2000) poses this question for general learning models and shows that introducing (bounded) memory can have quite strong implications.<sup>4</sup> For instance, where Kandori, Mailath, and Rob (1993) show the selection of risk dominant equilibria in coordination games, Alós-Ferrer (2000) shows that the

---

<sup>3</sup>The first versions of the results in Rhode and Stegeman (2001) were circulated in 1992.

<sup>4</sup>Young (1993) studies a different model where agents, sampled from two different populations, obtain a sample of records from past interactions before engaging in an asymmetric game. Due to the addition of sampling and fundamental differences in the modeling of agent interaction, Young’s model of memory can not be directly compared to the one in Alós-Ferrer (2000).

addition of memory leads to the selection of Pareto efficient equilibria. More importantly, it is shown that, even in models based on imitation and experimentation, where best response plays no role, there is still a certain significance to the Nash equilibrium, since the intertemporal comparison of payoffs after a unilateral deviation from a Nash equilibrium will always, by definition, allow the agents to perceive such deviation as an error and correct it by imitation of past (before-deviation) strategies.

This paper explores the impact of adding finite memory to the models of Cournot Oligopoly mentioned above. It is shown that, as long as memory includes at least one period, the walrasian outcome is no longer the only long-run outcome as in Vega-Redondo (1997). On the contrary, there is a clear tension between the walrasian outcome and the Cournot equilibrium, which stabilizes the whole range of quantities between them.

This result holds even with a *single* period of memory, for an arbitrary number of firms. This shows that the result in Vega-Redondo (1997), which is based on the relative payoff considerations mentioned above, is not robust with respect to the absolute payoff considerations introduced by memory. Just as in Vega-Redondo (1997), firms are pure imitators (i.e. they are neither supposed to know their profit functions nor to be able to compute a best reply). Hence, the difference in the results is due exclusively to the introduction of memory.

## 2 The Model

### 2.1 Preliminaries

Vega-Redondo (1997) studies a model where boundedly rational firms compete in quantities following a dynamic process of imitation and experimentation. The striking conclusion is that the Cournot-Nash equilibrium plays no role in the long-run, but rather the walrasian quantity is selected (i.e. a non-Nash outcome of the one-shot game).

In this paper, we examine the introduction of memory in such a framework. For tractability, we will work with a well-behaved Cournot Oligopoly, as reflected by the following technical assumptions.

There are  $N \geq 2$  firms producing a homogeneous good whose demand is given by the inverse-demand function  $P : \mathbb{R}_+ \mapsto \mathbb{R}_+$ . We assume this function to be twice-differentiable in a closed interval  $[0, Q_{max}]$ , downward-sloping (in the differentiable sense:  $P'(\cdot) < 0$ ) and concave ( $P''(\cdot) \leq 0$ ). Furthermore, we assume that  $P(0) = P_{max} > 0$  and  $P(x) = 0$  for all  $x \geq Q_{max}$ . All firms have an identical cost function  $C : \mathbb{R}_+ \mapsto \mathbb{R}_+$ , assumed upward-sloping ( $C'(x) > 0 \forall x > 0$ ) and differentially strictly convex ( $C''(\cdot) > 0$ ). A meaningful market requires the assumption that  $C'(0) < P_{max}$ .

**Definition 1.** The Walras quantity  $x^W$  is such that  $P(Nx^W)x^W - C(x^W) \geq P(Nx^W)x - C(x)$  for all  $x$ .

**Definition 2.** The Cournot quantity  $x^C$  is such that  $P(Nx^C)x^C - C(x^C) \geq P((N-1)x^C + x)x - C(x)$  for all  $x$ .

Under the assumptions above, it is straightforward to show that there exists a unique, strictly positive Cournot quantity  $x^C$  and also a unique, strictly positive Walras quantity  $x^W$ .

We also assume that all output levels that firms can announce belong to a common finite grid  $\Gamma = \{0, \delta, \dots, v\delta\}$ , with arbitrary  $\delta > 0$ , thought to be small, and  $v \in \mathbb{N}$ . Those quantities of interest, namely the Walras and Cournot ones, are assumed to be in the grid.

## 2.2 The model without memory

The dynamic model without memory corresponds to Vega-Redondo (1997). Time is discrete. Firms play the Cournot game at each time period, choosing output levels from  $\Gamma$ . They know neither the demand function nor the cost function. When choosing an output level for period  $t+1$ , the only information firms have available are the quantities produced and the profits realized by all firms in period  $t$ . A typical state of the process then includes  $N$  quantities  $x_1, \dots, x_N$  and the associated realized profits  $\pi_1, \dots, \pi_N$ . For simplicity, we shall work with a reduced state space as follows. Given the quantities  $\mathbf{x} = (x_1, \dots, x_N)$  produced in a given period, we can compute the profits obtained by each firm through

$$\pi_i = \Pi(x_i, x_{-i}) = P \left( \sum_{j=1}^N x_j \right) x_i - C(x_i).$$

where  $x_{-i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_N)$ . Formally, a typical state can be summarized by specifying the output levels only, i.e. the state space is simply  $\Gamma^N$ , in spite of the fact that firms observe quantities and realized profits. Note that this does not imply that firms are able to compute profits from quantities.

In the dynamics studied by Vega-Redondo (1997), firms observe the quantities produced and the profits realized by all competitors, and then imitate the quantities that led to the highest profits. Independently across time and firms, with a (small) probability  $\varepsilon \geq 0$ , a firm will instead experiment with a randomly chosen new quantity (all quantities in the grid having positive probability of being chosen). This event is called “a mutation.” The dynamics without mutations ( $\varepsilon = 0$ ) is called “unperturbed,” but the focus is on “perturbed” dynamics with  $\varepsilon > 0$ .

Fix a probability  $0 \leq \varepsilon < 1$ . Denote by  $x_i(t)$  the output level of firm  $i$  in period  $t$ , and by  $x_{-i}(t)$  the vector of output levels of its competitors that period. The behavior of each single firm  $i$  can be formally described as follows:

**Imitation** (occurs with probability  $1 - \varepsilon$ ). Let  $B_t$  be the set of observed output levels which yielded the highest profit in period  $t$ , i.e.

$$B_t = \{x_j(t) \mid j \in \{1, \dots, N\} \text{ and } \pi_j(t) \geq \pi_{j'}(t) \ \forall j' = 1, \dots, N\}$$

where  $\pi_{j'}(t) = \Pi(x_{j'}(t), x_{-j'}(t))$ . Then  $x_i(t+1) = x_j(t)$ , where  $x_j(t) \in B_t$ , selected randomly if  $B_t$  is not a singleton (according to a distribution with full support on  $B_t$ ). This behavior corresponds to an imitation rule which we will term “imitate the best.”

**Experimentation** (occurs with probability  $\varepsilon$ ).  $x_i(t+1)$  is some quantity chosen at random from  $\Gamma$ , according to a probability distribution with full support on  $\Gamma$ .

Imitation and experimentation happen independently across firms, i.e. several firms might experiment the same period but this is an unlikely event for small  $\varepsilon$  (the probability of any two given firms experimenting in a given period is  $\varepsilon^2$ ). Formally, the model described is a finite Markov chain (for each fixed  $\varepsilon$ ) which can be studied using the by-now standard techniques brought into economics by Kandori, Mailath, and Rob (1993) and Young (1993) (see Kandori and Rob (1995) or Ellison (2000) for other self-contained summaries). The objective is to find the *stochastically stable states*, which are those in the support of the limit invariant distribution of the process as the experimentation probability  $\varepsilon$  goes to zero. The interpretation is that, for small but positive  $\varepsilon$ , in the long run the process will spend most of the time on the stochastically stable states.

Given a quantity  $x \in \Gamma$ , denote  $\text{mon}(x) = (x, \dots, x) \in \Gamma^N$ . Such states are called *monomorphic*. The first, trivial observation is that the absorbing sets of the unperturbed dynamics ( $\varepsilon = 0$ ) correspond to singletons formed by monomorphic states (unsurprisingly, since the only force at work is imitation). It is then a standard result that only these states might be stochastically stable. Vega-Redondo (1997) shows that, in the dynamics described above, the only stochastically stable state is  $\text{mon}(x^W)$ , i.e. the walrasian equilibrium. The key to this result is the strong relative advantage that a firm producing  $x^W$  gets. Specifically, we can prove the following result. The proof is taken from Vega-Redondo (1997, p.381) and included here only for completeness.

**Lemma 1.** *For all  $x \neq x^W, 1 \leq m \leq N$ ,*

$$P((N-m)x + mx^W)x^W - C(x^W) > P((N-m)x + mx^W)x - C(x)$$

*Proof.* Since  $P$  is strictly decreasing,  $P(Nx^W) - P((N-m)x + mx^W)$  has the same sign as  $x - x^W$ . Hence,  $[P(Nx^W) - P((N-m)x + mx^W)] \cdot (x - x^W) > 0 \forall x \neq x^W$ , which amounts to  $P((N-m)x + mx^W) \cdot (x - x^W) < P(Nx^W) \cdot (x - x^W)$ . Adding  $C(x^W) - C(x)$  to both sides, we have

$$\begin{aligned} [P((N-m)x + mx^W)x - C(x)] - [P((N-m)x + mx^W)x^W - C(x^W)] < \\ [P(Nx^W)x - C(x)] - [P(Nx^W)x^W - C(x^W)] \end{aligned}$$

The right-hand side of the inequality is negative by definition of  $x^W$ , and hence the left-hand side is strictly negative, which proves the claim. ■

Taking  $m = N - 1$ , we see that the Walras quantity features *spite* (see e.g. Schaffer (1988) or Crawford (1991)): if a mutant deviates from the state  $\text{mon}(x^W)$ , even to a best-response, the payoffs of the non-mutants, still choosing  $x^W$ , will rise even more than the mutant's payoff.

Taking  $m = 1$ , we see that the Walras quantity also features what we could call *global stability* (see Alós-Ferrer (2000)): if a mutant firm deviates to  $x^W$  from any monomorphic state,  $\text{mon}(x)$ , with  $x \neq x^W$ , even if its payoffs after deviation have decreased, they will still be higher than those of the non-mutants. For instance, suppose the economy is at the Cournot equilibrium, i.e. the state  $\text{mon}(x^C)$ . If a mutant firm deviates to  $x^W$ , its profits will decrease, but the profits of the other firms will decrease even more. Hence, in absence of memory, the mutation is successful.

These two properties are enough to guarantee that  $\text{mon}(x^W)$  is the only stochastically stable state of the dynamics without memory. Intuitively, it is much more probable to reach this state from any other than to leave it.<sup>5</sup>

### 2.3 The model with memory

We want to investigate the effects of introducing of memory in this framework. Assume now that, when choosing output levels for next period, firms remember the quantities produced and the profits realized in the last  $K \geq 0$  periods in addition to the current one ( $K = 0$  is the model without memory). Formally, the state space is enlarged to  $\Gamma^{N \cdot (K+1)}$ , and the imitation rule simply specifies copying one of the quantities that led to highest payoffs in the last  $K + 1$  periods (including the present one).

The intuition is that, if firms remember past profits, destabilizing Cournot will not be such an easy task. After a single mutation from  $\text{mon}(x^C)$ , the mutant may earn more than the non-mutants, but the largest profit remembered will still be that of the Cournot equilibrium, and hence the mutant will “correct the mistake,” even in absence of any strategic considerations. This allows us to interpret the experimentation process as “trial-and-error”.<sup>6</sup>

Fix a probability  $0 \leq \varepsilon < 1$ . As in the model without memory, denote by  $x_i(t)$  the output level of firm  $i$  in period  $t$ , and by  $x_{-i}(t)$  the vector of output levels of its competitors that period. Formally, the behavior of each firm  $i$  is described as follows:

**Imitation** (occurs with probability  $1 - \varepsilon$ ). Let  $B_t^K$  be the set of observed output levels which yielded the highest profit in memory, i.e. in the last  $K + 1$  periods:

$$B_t^K = \{x_j(t - k) \mid j \in \{1, \dots, N\}, k \in \{0, \dots, K\} \text{ and} \\ \pi_j(t - k) \geq \pi_{j'}(t - k') \quad \forall j' = 1 \dots, N, k' = 0, \dots, K\}$$

where  $\pi_{j'}(t - k') = \Pi(x_{j'}(t - k'), x_{-j'}(t - k'))$ . Then  $x_i(t + 1) = x_j(t - k)$ , where  $x_j(t - k) \in B_t^K$ , selected randomly if  $B_t^K$  is not a singleton (according to a distribution with full support on  $B_t^K$ ). As in the no-memory case, we will refer to this imitation rule (which is discussed at length below) as “imitate the best.”

**Experimentation** (occurs with probability  $\varepsilon$ ).  $x_i(t + 1)$  is some quantity chosen at random from  $\Gamma$ , according to a probability distribution with full support on  $\Gamma$ .

Again, imitation and experimentation happen independently across firms, i.e. several firms might experiment the same period.

<sup>5</sup>In the terms of Kandori and Rob (1995), the minimum-cost  $\omega$ -tree is constructed connecting all states to  $\text{mon}(x^W)$  with a single mutation (global stability). Since it is impossible to leave  $\text{mon}(x^W)$  with a single mutation (spite), all the  $\omega$ -trees of other states have strictly larger costs. In the terms of Ellison (2000),  $\text{mon}(x^W)$  has Coradius equal to one by global stability, and Radius larger than or equal to two by spite.

<sup>6</sup>Vega-Redondo (1997) further assumes the presence of exogenous inertia, which has no effect there. This assumption, though, would effectively prevent the analysis of memory by enabling transitions where agents are “frozen” until the relevant payoffs are forgotten. See Alós-Ferrer (2000) for a discussion.

The particular case  $K = 0$  encompasses the model without memory. The general model is a Markov chain with state space  $\Gamma^{N \cdot (K+1)}$  which can be studied using the techniques mentioned above. In particular, given a quantity  $x \in \Gamma$ , denote  $\text{mon}(x, K) = ((x, \dots, x), \dots, (x, \dots, x))$ , i.e. the state where all firms have produced  $x$  for the last  $K + 1$  periods. In the model with memory, we call these states *monomorphic*. Obviously, the recurrent classes of the unperturbed process are the singletons formed by monomorphic states. In the absence of mutation, the imitation process can not lead away from a monomorphic state, and, given any non-monomorphic state, there is always strictly positive probability that all firms imitate the same quantity.<sup>7</sup> This guarantees that the analysis can be restricted to the monomorphic states. See e.g. Young (1993) or Ellison (2000).

## 2.4 The Imitation rule: a discussion

This paper focuses on a particular imitation rule, called “imitate the best.” This rule prescribes to imitate the strategy that yielded the highest profits in memory.

Ultimately, we think of firms’ decisions as being taken by human decision makers, e.g. managers. There seems to be no question that imitation of successful behavior is a common practice.<sup>8</sup> The particular imitation rule we have described, though, can be criticized on the grounds of its simplicity. Other, more sophisticated imitation rules spring to mind, as e.g. imitation of the strategy with the highest recalled average payoff, imitation of the highest within-period relative payoffs (Bergin and Bernhardt (1999)), or imitation based on the average population payoff (Oechssler (2002)), among many others. The position of this paper is that the imitate the best rule is worthy of attention, on the following grounds.

### 2.4.1 Salience of high payoffs

The described imitation rule is admittedly very naive. For instance, a single occurrence of a prominent high payoff might make agents favor a given strategy for  $K$  periods, although, if this strategy consistently yields low payoffs afterwards, that occurrence will eventually be forgotten. This apparent shortcoming is actually consistent with experimental evidence on human behavior. Barron and Erev (2003) and Erev and Barron (2003) report on a large number of decision-making experiments and identify several effects which lead to deviations from payoff maximization. The *Payoff Rank Effect* refers to the observation that alternatives with the highest recent payoffs are attractive to decision makers, even if they are associated with a low expected return. This illustrates that imitate the best might be more realistic than decision rules based on the imitation of

<sup>7</sup>With memory, the argument is slightly more complex. Each of the following  $K + 1$  periods, there is positive probability that all firms imitate the same quantity as other firms (though not necessarily the same as in other periods). This leads to a state of the form  $((x_1, \dots, x_1), (x_2, \dots, x_2), \dots, (x_{K+1}, \dots, x_{K+1}))$ . For the next  $K + 1$  periods all firms imitate the (remembered) quantity that yields higher profits when all firms produce it.

<sup>8</sup>Recently, Williams and Miller (2002) surveyed 1,684 executives to study their decision-making processes. A cluster analysis allowed for a classification of the executives in five groups. The most numerous group (36 %) was termed “followers”, who “make decisions based on how they’ve made similar decisions in the past or how other trusted executives have made them.”

average payoffs. The *Big Eyes Effect* shows a tendency to switch to strategies with the most attractive forgone payoffs (e.g. those that would have been obtained with unused strategies, or, in our framework, those actually obtained by other agents) in the recent trials. Both effects show a clear relationship to a known characteristic of human memory: *salient* outcomes (e.g. those leading to high payoffs) feature more prominently than others.

### 2.4.2 Simplicity

Another important reason to study the imitate the best rule is precisely its simplicity. It describes a particularly clear and straightforward way to take memory into account and allows us to obtain a tractable model where the main points of the paper can still be made and explained intuitively. However, we will show that the rule is rich enough to give interesting behavioral patterns.

Indeed, the combination of the imitate-the-best rule with memory and the possibility of experimentation allows for different learning considerations. First, since the relevant data includes all payoffs obtained (simultaneously) by agents in the most recent periods, the considerations arising from imitation models without memory (e.g. spite and the relevance of relative payoffs) are encompassed. For instance, suppose all firms are producing the Walrasian quantity. Suppose now a single firm experiments with a best response to this profile, which yields higher profits for the experimenter, but induces even larger payoffs for the remaining firms (by Lemma 1), still producing the Walras quantity. Although the experimenting firm has obtained higher profits than before, it is the remaining ones who have profited the most from the experiment, and hence firms will copy the Walras quantity, and not the deviation.

Second, the relevant data also includes the sequence of profits obtained by a given firm in different periods. As commented above, this allows for better-response and trial-and-error learning considerations. For instance, suppose a single firm deviates from the Cournot-Nash equilibrium through an experiment. Even though, after deviation, this firm might earn more than the remaining firms, it will certainly obtain lower payoffs than in the previous period, and then will be able to recognize the experiment as a mistake (trial and error). This and the previous observation, though, would be obscured under more sophisticated learning rules. Consider, e.g., the “imitate the best average” rule. Suppose firms are producing the Cournot-Nash equilibrium quantities. Under the imitate-the-best-average rule, agents would only recognize a deviation from this Nash equilibrium as a mistake if the deviation payoff were lower than the (weighted) average of the Cournot equilibrium profit and the profit yielded by the Cournot quantity in the post-deviation profile.

### 2.4.3 Low decision costs

The imitate the best rule poses a minimal computational burden on the side of agents (firms), while still incorporating memory. Conlisk (1980) and Pingle and Day (1996), among others, pointed out that imitation plays an important role in real-world economic problems because it allows agents to economize on decision costs.<sup>9</sup> Under the imitate the best rule, agents do not need to store

---

<sup>9</sup>Pingle and Day (1996) remark that both imitation and trial-and-error learning can be seen as ways to economize on decision costs.

processed information, such as weighted averages or estimated beliefs. Indeed, to implement this rule, it is enough to remember a single payoff and strategy per period and make sequential comparisons with observed data (this fact will help simplify the analysis). As a consequence, the imitate the best rule could still be studied in a framework where firms' attention is focused on the industry's top performers each period ("benchmarking", "best practices", etc).

#### 2.4.4 Relationship to the literature

Sarin (2000) develops an alternative model of memory in a non-strategic framework (a multi-armed bandit). Rather than remembering the actions and payoffs of the latest  $K$  periods, the agents in Sarin (2000) remember the latest  $K$  payoffs attained by every single available action. For instance, an agent might remember the payoffs obtained in the latest three periods by an action which has effectively been used, and three payoffs that another, long-discarded action yielded one hundred periods ago. With this *caveat*, it must be noted that the behavioral rule studied in this paper and in Alós-Ferrer (2000) is analogous to the Max function in Sarin (2000, p.155), which is presented there as a model for "optimistic" agents.

The very idea of trial-and-error learning implies, at least, one period of memory. Intuitively, an agent learning by trial and error should (randomly) sample a new strategy, try it out and compare the results with previous ones. This implies the capability of remembering previous outcomes. This is precisely the view taken by Gale and Rosenthal (1999), who analyze a model with a population of imitators and a single experimenter. The experimenter tries out a random new strategy, and only then decides whether to keep it or to revert to the previous one. In contrast to the present paper, experimentation in Gale and Rosenthal (1999) is virtual, in the sense that, although the experimenter can observe the payoffs of the tried strategy against a static profile, the trial does not actually take place. Since the new strategy is only adopted if successful, this effectively gives rise to "better-reply" behavior.

The trial-and-error interpretation of the imitation and experimentation process might be used to relate this paper to the reinforcement learning literature. Successful strategies, identified through intertemporal comparisons, are likely to be reinforced, i.e. used again.

This statement needs to be clarified. Probabilistic reinforcement learning models, in the sense used in psychology and recent economics literature, do not support payoff maximization (due to "probability matching" phenomena) or Nash outcomes.<sup>10</sup> However, the computer science literature often considers reinforcement learning models based on trial-and-error experimentation (for an introduction, see Kaelbling, Littman, and Moore (1996) or Sutton and Barto (1998)). Clearly, these models are conceptually related to the trial-and-error part of the model presented here. They are conceived as optimization techniques ("machine learning"), hence aiming at Nash equilibria in games (provided agents do not revise strategies simultaneously).

<sup>10</sup>Erev and Roth (1998) show that experimental data on play in games with unique, mixed Nash equilibria can be explained through a model of reinforcement learning, but actual play moves away from equilibrium. On a related note, although their formal model is rather different to the one considered here, they also show that its descriptive power increases when two additional parameters are introduced: "experimentation" and "forgetting."

### 3 Analysis

#### 3.1 Relative advantage vs. absolute success

Two important comparisons capture the relative advantages of any given quantity. The first is the difference in payoffs between a mutant and the non-mutants after a single mutation from a monomorphic state. This is given by:

$$D(x, y) = P((N - 1)x + y)(y - x) + C(x) - C(y)$$

i.e. the difference, after mutation, between the payoffs of a mutant producing  $y$  and the non-mutants, all of them still producing  $x$ . If  $D(x, y) > 0$ , then, in a model without memory, a single mutation is enough for a transition from  $\text{mon}(x)$  to  $\text{mon}(y)$ . The spite of  $x^W$  translates to  $D(x^W, y) < 0$  for all  $y \neq x^W$ , and the global stability amounts to  $D(x, x^W) > 0$  for all  $x \neq x^W$ . Both facts, which follow from Lemma 1, are illustrated in Figure 1.

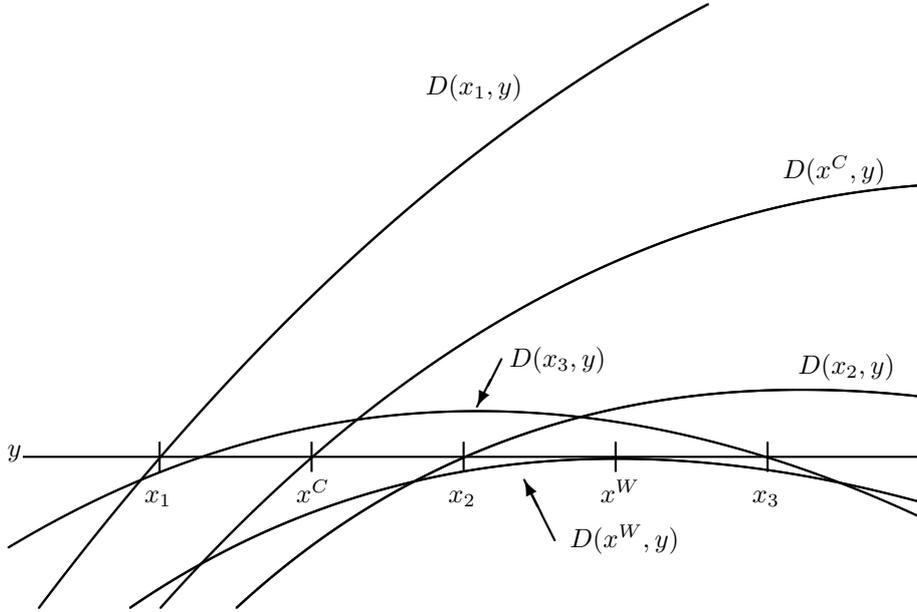


Figure 1:  $D(x, y)$  is the payoff difference between a mutant playing  $y$  and the non-mutants playing  $x$ . If it is positive, the mutant fares better. The horizontal axis is the second variable of  $D(\cdot, \cdot)$ , i.e. the mutation  $y$ .

The second important comparison is the difference between the payoff of a mutant, producing  $y$  after mutation (from a monomorphic state where all firms produce  $x$ ) and the payoff that the same mutant had before mutation, that is, the absolute gain or loss due to the mutation:

$$M(x, y) = P((N - 1)x + y)y - C(y) - P(N \cdot x)x + C(x)$$

If  $M(x, y) > 0$ , then the mutant obtains, after deviation, higher profits than before. The fact that the Cournot outcome is a (strict) Nash equilibrium translates to  $M(x^C, x) < 0$  for all  $x \neq x^C$ .

It is shown in Lemma A.3 (see Appendix A) that, under our assumptions,  $M(x, x^C) > 0$  for all  $x \neq x^C$ , which shows a certain analogy between the Walras and Cournot quantities. The Walras quantity is such that deviations to or from it always render a situation where relative payoffs are higher for the firms producing  $x^W$  after mutation. The Cournot quantity is such that single deviations to or from it always render a situation where an “intertemporal” comparison between pre- and post-mutation payoffs for the mutant always favors  $x^C$ . This is illustrated in Figure 2.

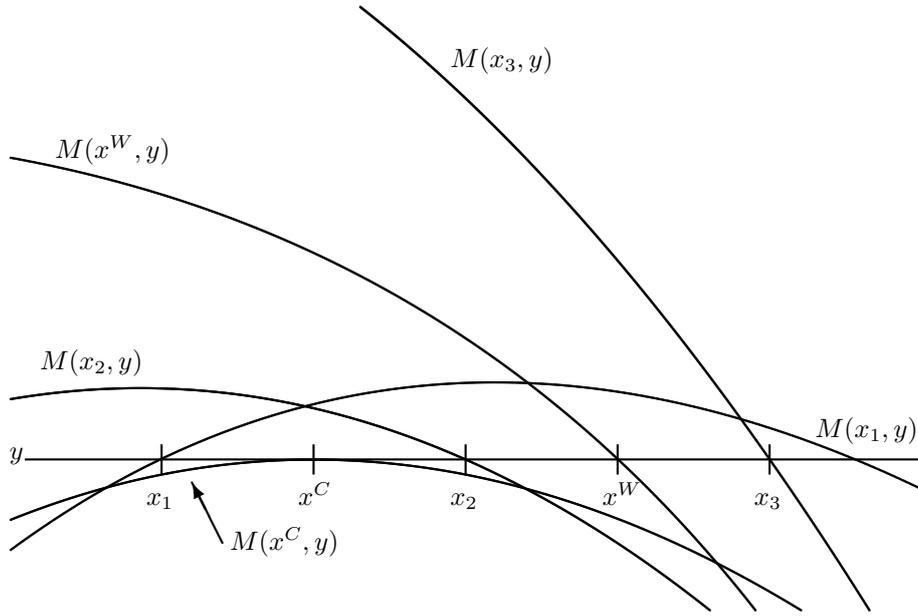


Figure 2:  $M(x, y)$  is the payoff difference experienced by a mutant deviating from  $x$  to  $y$ , when all other firms keep producing  $x$ . If it is positive, the mutant is better off. The horizontal axis is the second variable of  $M(\cdot, \cdot)$ , i.e. the mutation  $y$ .

Even though, potentially, each firm might have a different output level each period, the most relevant quantities for the analysis of the model are  $D(x, y)$  and  $M(x, y)$ , which only depend on two output levels. This is because, since we are dealing with an imitation model, attention can be restricted to monomorphic states and deviations from them. Since the imitate the best rule does not require processed information, to evaluate a deviation  $y$  from a monomorphic state  $\text{mon}(x, K)$ , only the quantities  $x$  and  $y$  are required.

The quantities  $D(x, y)$  and  $M(x, y)$  are not unrelated. Lemma A.1 (see Appendix A) shows that a deviation to a lower quantity ( $y < x$ ) is always better in absolute than in relative terms (i.e.  $M(x, y) > D(x, y)$ ), whereas exactly the opposite is true for deviations to higher quantities.

In a model with memory, both absolute and relative payoff comparisons are crucial. In the remaining analysis we see how this observation yields *both* the Walras and the Cournot quantities as stable outcomes.

### 3.2 The stable set: quantities between Cournot and Walras

As a first approximation, the following theorem shows that both quantities above  $x^W$  and below  $x^C$  can be quickly discarded.

**Theorem 1.** *For any  $K \geq 1$ , the set of stochastically stable states is contained in*

$$\{\text{mon}(x, K)/x^C \leq x \leq x^W, x \in \Gamma\}$$

*i.e. it is a subset of the set of monomorphic states corresponding to quantities between (and including) the Cournot and the Walras ones. Moreover,  $\text{mon}(x^W)$  is always stochastically stable.*

The proof (which is relegated to Appendix A) proceeds as follows. From a monomorphic state, single mutations leave the mutant with higher after-mutation profits than those of the non-mutants only if the mutation happens in the direction of  $x^W$  (see Figure 1). Such mutations are thus successful in relative terms, but not necessarily in absolute terms. Analogously, single mutations leave the mutant with higher profits than before only if the mutation happens in the direction of  $x^C$  (see Figure 2). Such mutations are thus successful in absolute terms, but not necessarily in relative ones.

This is illustrated in Figure 3, which depicts  $M(x, x^W)$ ,  $M(x, x^C)$ ,  $D(x, x^W)$ , and  $D(x, x^C)$ . Whereas the walrasian quantity is guaranteed to represent an advantage in relative, spiteful terms (due to Lemma 1), the Cournot quantity is guaranteed to represent an advantage in absolute, intertemporal terms.

Quantities below  $x^C$  are unstable, because a mutation to a higher quantity (e.g.  $x^C$ ) can go simultaneously in the direction of  $x^C$  and  $x^W$ . A single mutant, hence, will earn higher payoffs than the non-mutants both after and before mutation, and will be imitated. The same holds for quantities above  $x^W$ , because a mutation to a lower quantity (e.g.  $x^W$ ) can go simultaneously in the direction of  $x^W$  and  $x^C$ .

Quantities in  $[x^C, x^W]$ , though, are more stable. More precisely, no single mutation from the corresponding monomorphic state can be successful. This fact is established in Lemma A.6. The argument (which is partially reflected in Figure 3), is as follows.

Consider any quantity  $x \in [x^C, x^W]$ . A mutation from  $\text{mon}(x, K)$  “upwards,” towards a higher quantity (e.g.  $x^W$ ) might leave the mutant with higher after-mutation profits than the non-mutants ( $D(x, x^W) > 0$ , see Figure 3). Still, before-mutation profits are the highest profits remembered ( $M(x, x^W) < 0$ ). Hence, the mutation would not be successful. A mutation “downwards,” towards a lower quantity (e.g.  $x^C$ ) might leave the mutant with higher payoffs than before mutation ( $M(x, x^C) > 0$ ). The largest payoff observed, though, would be that of the non-mutants (after mutation) ( $D(x, x^C) < 0$ ). Hence, such a mutation would also be unsuccessful. In summary, a single mutation from a state where all firms produce  $x \in [x^C, x^W]$  will never be successful (Lemma A.6), whereas a single mutation is enough to destabilize all other states (Lemmata A.4 and A.5). This reflects a different level of stability and is enough to drive the result: all stochastically stable states must be monomorphic states corresponding to quantities between the Cournot and the Walras ones.<sup>11</sup>

<sup>11</sup>These arguments imply that, in the terminology of Ellison (2000), the set

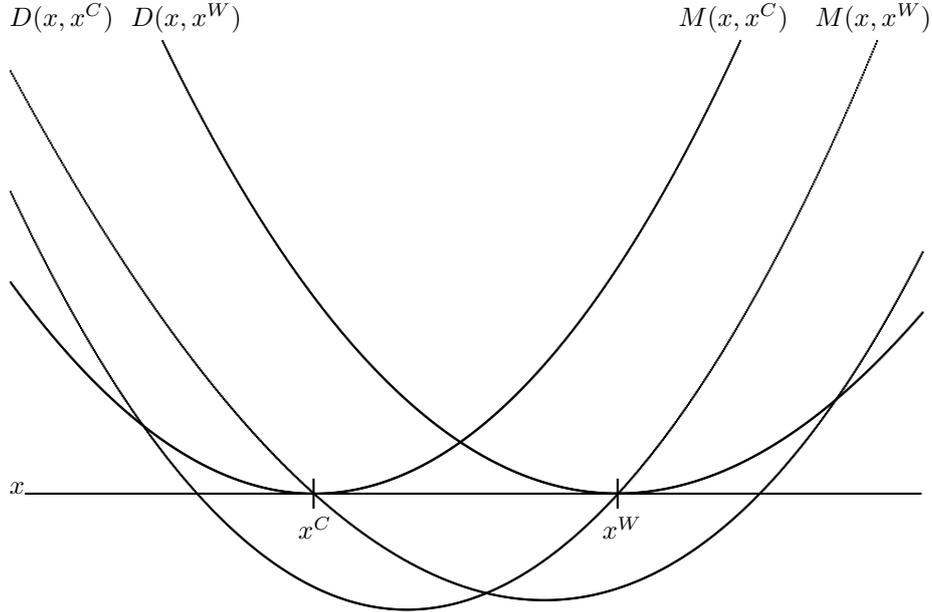


Figure 3: Deviations to Cournot and Walras. A mutant from  $x$  to  $y = x^C$  or  $x^W$  gains in relative terms if and only if  $D(x, y) > 0$ . He gains in absolute terms if and only if  $M(x, y) > 0$ . The horizontal axis is here the first variable of the functions  $D(\cdot, \cdot)$  and  $M(\cdot, \cdot)$ .

Theorem 1 also states that the walrasian quantity is always stochastically stable. The essence of the proof is the following. Starting from a quantity  $x \in [x^C, x^W)$ , it is always possible to move upwards from  $\text{mon}(x, K)$  to  $\text{mon}(x + \delta, K)$ , where  $x < x + \delta \leq x^W$ , with two simultaneous mutations. For example, from  $\text{mon}(x, K)$ , consider two simultaneous mutations to quantities  $x - \delta$  and  $x + \delta$ . This leaves the total produced quantity, and hence the market price  $p = P(N \cdot X)$ , unchanged. With a fixed price, though, the profit function  $p \cdot x - C(x)$  is concave and attains its maximum above  $x^W$  (see Lemma A.7). Hence, the mutant producing  $x + \delta$  earns larger profits than the non-mutants (both before and after mutation), and than the mutant producing  $x - \delta$ . Thus,  $x + \delta$  is imitated.

Repeating this argument, we can construct a chain that connects  $\text{mon}(x, K)$  to  $\text{mon}(x^W, K)$ , where each transition is done with exactly two mutations. In the terminology of Kandori, Mailath, and Rob (1993), this is enough to build the  $\text{mon}(x^W, K)$ -tree of minimal cost.

$\overline{\{\text{mon}(x, K)/x^C \leq x \leq x^W, x \in \Gamma\}}$  has Radius strictly larger than one, but Coradius 1. Theorem 1 in Ellison (2000) then yields an alternative proof of the result, and additionally implies that the estimated time of first arrival into the mentioned set is of order  $\varepsilon^{-1}$ , that is, convergence is as fast as it can be, and independent of population size.

### 3.3 The duopoly case

Theorem 1 shows that the Cournot quantity and the Walras one are respectively the lower and upper bound of the quantities that can be observed in the market in the long run. Quantities that are too low (respectively too high) are promptly destabilized because deviations to higher (respectively lower) quantities are made both in the direction of the Cournot quantity and the Walras one. Hence, the result makes use of two powerful forces. The first force is the one associated to Nash equilibria, reinterpreted as an intertemporal comparison of own payoffs. The second, the effects of spite and global stability.

Between  $x^C$  and  $x^W$ , these two forces clash. We see now a first case where the outcome is “undecided” and their clash stabilizes the whole interval of quantities: the duopoly.

**Proposition 2.** *If  $N = 2$  (a duopoly), and for any  $K \geq 1$ , the set of stochastically stable states is identically equal to*

$$\{\text{mon}(x, K)/x^C \leq x \leq x^W, x \in \Gamma\}$$

*i.e. it is formed by all the monomorphic states corresponding to quantities between (and including) the Cournot and the Walras ones.*

The proof is relegated to Appendix A. The key idea is that, with a population of just two firms, after two simultaneous mutations to the same quantity there are no relative-payoff considerations, because there are no non-mutants left. If only absolute-payoff considerations are present, mutations in the direction of  $x^C$  are again successful. This allows to construct transitions “downwards” from monomorphic states corresponding to quantities  $x \in (x^C, x^W]$  with two mutations. This implies that, within  $[x^C, x^W]$ , it is just as costly (in terms of mutations) to move up or downwards. Upwards transitions are as mentioned above for the  $\text{mon}(x^W, K)$ -tree of minimal cost. Downwards transitions use two simultaneous mutations to the same (smaller) quantity.

Thus, although the walrasian quantity is always contained in the prediction, it will in general not be the only quantity there, as in Vega-Redondo (1997). Moreover, the Cournot quantity might be in the prediction.

### 3.4 The general case with more than two firms

We concentrate now on  $N \geq 3$  (since the case  $N = 2$  is solved by Proposition 2). We will conclude that the whole interval  $[x^C, x^W]$  is also stable in this case, although the required construction to prove this result is not as simple as in the duopoly case.

By Theorem 1, the walrasian outcome is stochastically stable, and the only other candidates for stochastic stability are the states  $\text{mon}(x, K)$  with  $x \in [x^C, x^W]$ . The key to establish that all these states are indeed stable is that it is possible to destabilize Walras with the same cost (in terms of mutations) as any other quantity in  $[x^C, x^W]$ .

We need to consider transitions leaving  $\text{mon}(x^W, K)$ . This cannot be accomplished with one single mutation, but it is easy to see that two mutations would suffice. To get an intuition of how, we consider an extreme example. Let  $x_{N-1}^W > x^W$  be the Walras quantity when there are only  $N - 1$  firms in the market. Suppose two mutant firms produce respectively 0 and  $x_{N-1}^W$ . In

all practical respects, the situation is as if there were only  $N - 1$  firms in the market, and the one that now produces  $x_{N-1}^W$  earns larger profits than the rest by Lemma 1. But these firms earn more than before mutation, since their costs remain the same and the price has risen ( $x_{N-1}^W$  is easily seen to be lower than  $2 \cdot x^W$ ). In summary, the firm that has deviated to  $x_{N-1}^W$  is better off both in relative and absolute terms, and hence the transition to  $\text{mon}(x_{N-1}^W, K)$  follows.

Once a transition to a larger quantity has been achieved with two mutations, the process is in the unstable region out of  $[x^C, x^W]$ . The new monomorphic state can be destabilized by a single mutation to a lower quantity, and it can be computed that quantities strictly lower than  $x^W$  can be reached now in such a way. Consider one such quantity,  $x'$  with  $x^C \leq x' < x^W$ . This shows that  $\text{mon}(x', K)$  must be stochastically stable. Consider the  $\text{mon}(x^W, K)$ -tree which shows stability of Walras. Add the two transitions discussed above, with a total cost of three. In exchange, the previous transitions leaving  $\text{mon}(x_{N-1}^W)$  (at cost one) and  $\text{mon}(x', K)$  (at cost two) can be deleted, yielding for the new tree (with vertex  $\text{mon}(x', K)$ ) the same cost as that of the original  $\text{mon}(x^W, K)$ -tree.

Transitions such as these can be used to show stability of the whole spectrum of quantities in  $[x^C, x^W]$ , as the next theorem states. The detailed proof is relegated to Appendix B. Note that it suffices to add a single period of memory to alter the conclusions of the model without memory.

**Theorem 2.** *For any  $K \geq 1$ ,  $N > 2$ , and  $\delta$  small enough, the set of stochastically stable states is*

$$\{\text{mon}(x, K) / x^C \leq x \leq x^W, x \in \Gamma\}$$

*i.e. all the monomorphic states corresponding to quantities between the Cournot and the Walras ones.*

An example of the procedure used to show stability of non-walrasian quantities is illustrated in Figure 4 (based on quadratic costs and linear demand). At the state  $\text{mon}(x^W, K)$ , consider two mutations to quantities 0 and  $x$  with  $x^W < x < 2 \cdot x$ . Let  $H(x)$  be the after-mutation payoff difference between the  $x$ -mutant and the incumbents. As long as  $H(x)$  is positive, the  $x$ -mutant is successful. The function  $H(x)$  is easily seen to be strictly concave, which yields a range of monomorphic states, corresponding to quantities above  $x^W$ , which can be reached from  $\text{mon}(x^W, K)$  with just two mutations. But from each of these states, a single mutation to lower quantities  $y$  is successful as long as  $D(x, y) \geq 0$ . If it happens, for instance, that  $D(x, x^C)$  is positive, the stability of  $\text{mon}(x^C, K)$  follows.

## 4 Discussion

Memory allows agents to behave as if they were able to “experiment conditionally.” When a new strategy is tried out, memory allows to compare its success with that of the previous strategy. If the mutation brings payoffs down, the mutant will be able to “correct” the mistake and go back to the previous action. This observation, which is of intertemporal nature in an explicitly dynamic framework, naturally reintroduces better-response considerations into models of bounded rationality without explicitly assuming that the agents compute any

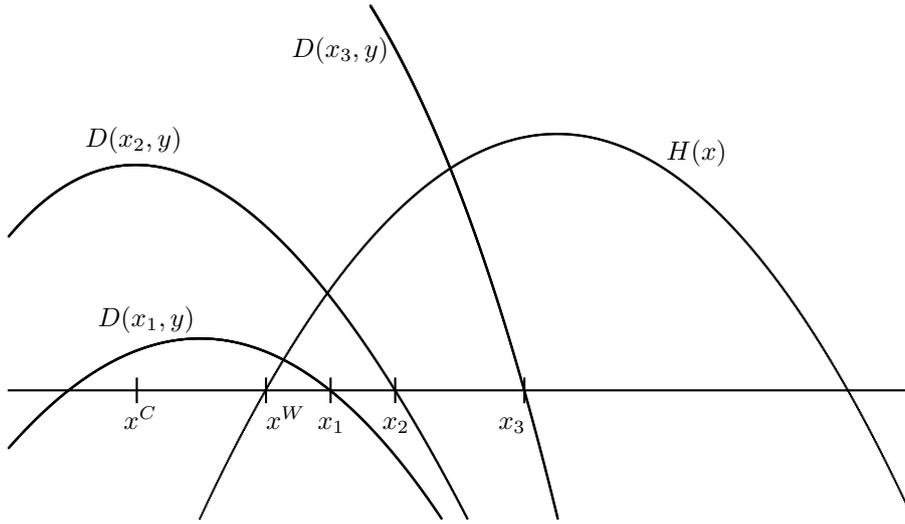


Figure 4: How to leave Walras. A transition from  $\text{mon}(x^W, K)$  to  $\text{mon}(x, K)$ ,  $x > x^W$  can be achieved with two mutations if and only if  $H(x) \geq 0$ . A transition from  $\text{mon}(x_i, K)$  to  $\text{mon}(y, K)$  (e.g.  $y = x^C$ ) can be achieved with one mutation if  $D(x_i, y) \geq 0$ . The horizontal axis is the second variable of  $D(\cdot, \cdot)$  and the argument of  $H(\cdot)$ .

best reply. However, the example analyzed in this paper shows that this is not the only consideration. It is the interplay between better response (to do better than yesterday) and relative success (to do better than the others) which creates a rich dynamic in which two properties determine the long-run outcomes. The first property is the one associated to Nash equilibria, reinterpreted as an intertemporal comparison of own payoffs. The second, the effects of spite and global stability.

In the case of a Cournot Oligopoly, we see the two forces clash, giving us an economically meaningful example where two focal points are selected: the first thanks to its Nash-equilibrium condition, the second thanks to its spiteful properties. As these two “forces” compete, the whole interval between those two outcomes is stabilized, creating an apparently blunt prediction. In actual examples, the prediction reduces to a small interval of quantities (necessarily shrinking for large number of firms), not unlike the interval of prices found in Alós-Ferrer, Ania, and Schenk-Hoppé (2000) for the Bertrand model. However, the fact that it is still an interval appropriately illustrates both the stability properties of the Cournot-Nash equilibrium and of the spite-driven walrasian outcome.

From the point of view of Industrial Organization, this paper shows the lack of stability of quantities above the Walras one, and below the Cournot one. It shows that both of these quantities play a clear role, limiting the interval of stable quantities, but neither of them is to be expected as a unique solution. Theorem 1 shows that, in any given path of the considered process, we would rarely observe quantities outside the interval limited by the Cournot and Walras quantities. Theorem 2 shows that, in the long run, all quantities within this

interval should be observed.<sup>12</sup>

The proof of Theorem 2 essentially identifies a set of high-probability paths which correspond to a dynamic, almost-cyclical behavior. In these paths, and once the market settles in a given quantity, eventually some firms deviate from it, both to lower and higher quantities, but only the ones with higher quantities are successful, due to spite-driven considerations. This process keeps raising the market quantity until some firm raises it too much, above the walrasian one, enjoying a short-lived prosperity which is quickly undermined by other firms switching to much lower quantities. From these lower quantities, the “cycle” can start again.

## A Proof of Theorem 1

The proof proceeds through a series of lemmata, some of which will also be useful for the proof of Theorem 2. Proofs of Lemmata A.2 and A.7 (and part of Lemma A.6) rely only on differential calculus and can be found in the journal’s webpage.

First, recall that the recurrent classes of the unperturbed process are the monomorphic states. Thus, we can concentrate the analysis on them (see e.g. Young (1993) or Ellison (2000)).

**Lemma A.1.**  $M(x, y) > D(x, y) \forall y < x; M(x, y) < D(x, y) \forall y > x$

*Proof.* Notice that  $M(x, y) = D(x, y) + x \cdot [P((N - 1)x + y) - P(N \cdot x)]$ . If  $y < x$ , then  $(N - 1)x + y < N \cdot x$ . Since  $P$  is strictly decreasing, this implies that  $P((N - 1)x + y) - P(N \cdot x) > 0$ , proving the claim. If  $y > x$ , the proof proceeds analogously. ■

**Lemma A.2.** *For any  $x$ ,  $M(x, \cdot)$  is a strictly concave function. Define  $y^M(x) = \operatorname{argmax}\{M(x, y) / 0 \leq y \leq Q_{\max}\}$ . Then,  $y^M(x)$  is continuous and decreasing. Moreover, there exists  $Q^M > x^C$  such that  $y^M(x)$  is differentiable and strictly decreasing in  $(0, Q^M)$  and  $y^M(x) = 0$  for all  $x \in [Q^M, Q_{\max}]$ .*

**Lemma A.3.**  $M(x, x^C) > 0$  for all  $x \neq x^C$

*Proof.* Notice that  $y^M(x^C) = x^C$ . From Lemma A.2, it follows that, for all  $x < x^C$ ,  $y^M(x) > x^C > x$ . Analogously, for all  $x > x^C$ ,  $y^M(x) < x^C < x$ . In both cases, since  $M(x, x) = 0$ ,  $M(x, \cdot)$  is strictly concave, and  $y^M(x)$  is its maximum, we have that  $M(x, x^C) > 0$ . ■

**Lemma A.4.** *For any  $x < x^C$ , a single mutation suffices for the transition from  $\operatorname{mon}(x, K)$  to  $\operatorname{mon}(x^C, K)$ .*

*Proof.* Starting at  $\operatorname{mon}(x, K)$ , suppose one mutant firm deviates to  $x^C$ . By Lemma A.3,  $M(x, x^C) > 0$ ; i.e. the mutant attains higher payoffs than all firms before mutation, when all produced  $x$ . By Lemma A.1,  $D(x, x^C) > M(x, x^C) >$

<sup>12</sup>Interestingly enough, the same prediction arises in some oligopoly models with highly sophisticated firms. In Ferreira (2003), any outcome between perfect competition and Cournot can be sustained as an equilibrium when firms have access to infinitely many futures markets prior to one spot market. Delgado and Moreno (2002) show that, if firms compete by submitting (increasing) supply functions, the set of equilibrium prices is the interval between the competitive price and the Cournot price.

0; i.e. the mutant also attains higher payoffs than the non-mutants after mutation. Since the mutant's is the highest observed payoff, the mutation is successful and the transition to  $\text{mon}(x^C, K)$  is completed after  $K$  consecutive periods. ■

**Lemma A.5.** *For any  $x > x^W$ , a single mutation suffices for the transition from  $\text{mon}(x, K)$  to  $\text{mon}(x^W, K)$ .*

*Proof.* Starting at  $\text{mon}(x, K)$ , suppose one mutant firm deviates to  $x^W$ . By Lemma 1,  $D(x, x^W) > 0$ ; i.e. the mutant attains higher payoffs than the non-mutants after mutation. By Lemma A.1,  $M(x, x^W) > D(x, x^W) > 0$ ; i.e. the mutant also attains higher payoffs than all firms before mutation, when all produced  $x$ . Since the mutant's is the highest observed payoff, the mutation is successful and the transition to  $\text{mon}(x^C, K)$  is completed after  $K$  consecutive periods. ■

**Lemma A.6.** *No state  $\text{mon}(x, K)$  with  $x^C \leq x \leq x^W$  can be destabilized with only one mutation.*

*Proof.* Consider a mutation to  $y > x$ . By Lemma A.2,  $M(x, \cdot)$  is strictly concave. Moreover, since  $x \geq x^C$ , it follows that  $y^M(x) \leq y^M(x^C) = x^C \leq x$ , and hence  $M(x, \cdot)$  is decreasing in  $[x, Q_{max}]$ . In particular,  $M(x, y) < M(x, x) = 0$ , i.e., after a single mutation to  $y > x$ , the mutant will obtain smaller profits than before. It follows that the process will simply go back to  $\text{mon}(x, K)$  in absence of further mutations.

Consider now a mutation to  $y < x$ . It is not difficult to show (see the journal's webpage for details) that  $D(x, \cdot)$  is increasing in  $[0, x]$ . Then,  $D(x, y) < D(x, x) = 0$  for  $y < x$ , i.e. the mutant obtains smaller profits than the non-mutants after a single mutation to  $y < x$ . It follows that the process will go back to  $\text{mon}(x, K)$  in absence of further mutations. ■

**Lemma A.7.** *For  $p \in [0, P_{max}]$ , the function  $g_p(z) = p \cdot z - C(z)$  is strictly concave in  $z$ , and it attains a unique maximum  $z(p)$ . The function  $z(p)$  is increasing and  $z(P(N \cdot x^W)) = x^W$ .*

**Lemma A.8.** *For all  $x^C \leq x \leq x^W - \delta$ , two mutations suffice for the transition from  $\text{mon}(x, K)$  to  $\text{mon}(x + \delta, K)$  to occur.*

*Proof.* Consider the situation  $\text{mon}(x, K)$ , where all firms are producing a quantity  $x$  and selling it at a price  $P = P(N \cdot x)$ . Consider two simultaneous mutations of the following type. One firm deviates to  $x + \delta$ , and another firm deviates to  $x - \delta$ . The market price in the new situation,  $(x + \delta, x - \delta, x, \dots, x)$ , is  $P(x + \delta + x - \delta + (N - 2) \cdot x) = P(N \cdot x) = P$ .

Since  $z(p)$  is increasing by Lemma A.7,  $P = P(N \cdot x) > P(N \cdot x^W)$  implies that  $z(P) \geq z(P(N \cdot x^W)) = x^W$ . Since the function  $g_P(z)$  is strictly concave and has a maximum at  $z(P)$ , this implies that it is strictly increasing on  $[0, x^W]$ , and, in particular,  $P \cdot (x + \delta) - C(x + \delta) > P \cdot (x) - C(x) > P \cdot (x - \delta) - C(x - \delta)$ . Hence, after the two mutations described above, the mutant producing  $x + \delta$  not only earns higher profits than the non-mutants, but also higher ones than any profits in the previous periods, implying that the transition to  $\text{mon}(x + \delta, K)$  will be successfully completed. ■

*Proof of Theorem 1.* We have to show that the monomorphic state  $\text{mon}(x^W, K)$  is stochastically stable, and that no monomorphic state  $\text{mon}(x, K)$  with  $x \notin [x^C, x^W]$  can be stochastically stable.

Call  $n_1$  the number of quantities in  $[x^C, x^W] \cap \Gamma$ , and  $n_2$  the number of quantities in  $\Gamma \setminus [x^C, x^W]$ .

Consider any  $\omega$ -tree. Stochastically stable states are those with minimal-cost  $\omega$ -trees (see e.g. Kandori, Mailath, and Rob (1993) or Ellison (2000)). In any such tree, the cost necessary to leave any state but the vertex is two for monomorphic states corresponding to quantities in  $[x^C, x^W] \cap \Gamma$  (by Lemmata A.6 and A.8), and at least one for all others (because monomorphic states can not be destabilized without mutation). If we are able to construct an  $\omega$ -tree such that

- (i) the vertex  $\omega$  corresponds to a state  $\text{mon}(z, K)$  with  $z \in [x^C, x^W] \cap \Gamma$ ,
- (ii) for all other  $x \in [x^C, x^W] \cap \Gamma$ , the state  $\text{mon}(x, K)$  is connected at cost two, and
- (iii) for all  $x \in \Gamma \setminus [x^C, x^W]$ , the state  $\text{mon}(x, K)$  is connected at cost one,

then this would show that the state  $\omega$  is stochastically stable. The cost of this tree would be  $2 \cdot (n_1 - 1) + n_2 = 2 \cdot n_1 - 2 + n_2$ . Moreover, it follows that the stochastically stable states are exactly those having  $\omega$ -trees of this cost.

We construct now such a tree with vertex  $\text{mon}(x^W, K)$ .

For all  $x < x^C, x \in \Gamma$ , the state  $\text{mon}(x, K)$  is connected at cost one to  $\text{mon}(x^C, K)$ , by Lemma A.4.

For all  $x > x^W, x \in \Gamma$ , the state  $\text{mon}(x, K)$  is connected at cost one to  $\text{mon}(x^W, K)$ , by Lemma A.5.

For all  $x \in [x^C, x^W] \cap \Gamma$ , the state  $\text{mon}(x, K)$  is connected at cost two to  $\text{mon}(x + \delta, K)$ .

This completes a  $\text{mon}(x^W, K)$ -tree of the desired cost. Hence,  $\text{mon}(x^W, K)$  is stochastically stable.

Consider any quantity  $x \in \Gamma \setminus [x^C, x^W]$ . The minimum cost that could be attained by any  $\text{mon}(x, K)$ -tree is at least  $2 \cdot n_1 + n_2 - 1$  by Lemma A.6, which is larger than required. This proves that  $\text{mon}(x, K)$  is not stochastically stable. ■

*Proof of Proposition 2.* If  $N = 2$ , then the state  $\text{mon}(x, K)$  can be connected to the state  $\text{mon}(x - \delta, K)$  with two mutations (both to  $x - \delta$ ) for all  $x > x^C, x \in \Gamma$ . This follows from the fact that the function  $P(2 \cdot x) \cdot x - C(x)$  is strictly concave and its maximum, that is, the collusion output, is lower than  $x^C$ .

For any  $x \in [x^C, x^W] \cap \Gamma$ , consider the  $\text{mon}(x^W, K)$ -tree constructed in the proof of the Theorem 1. Consider all the arrows leaving states  $\text{mon}(x', K)$  for  $x' \in [x, x^W] \cap \Gamma$ . These arrows, which connected  $\text{mon}(x', K)$  to  $\text{mon}(x' + \delta, K)$  with two mutations, can be reversed so that they connect now  $\text{mon}(x' + \delta, K)$  to  $\text{mon}(x', K)$ , also with two mutations. This procedure yields a  $\text{mon}(x, K)$ -tree of minimal cost, which, together with Theorem 1, proves the result. ■

## B Proof of Theorem 2

In this section, let  $N \geq 3$ . The proof of the main Theorem consists of a series of interrelated Lemmata. Lemmata B.3 to B.7 are of technical nature and their

often lengthy proofs (which involve only differential calculus) are omitted here. They are available in the webpage of the journal.

We already know that the stochastically stable states must be contained in the set of monomorphic states corresponding to quantities in  $[x^C, x^W]$ , and that  $\text{mon}(x^W, K)$  is stochastically stable. To prove Theorem 2, we have to build  $\text{mon}(x, K)$ -trees (with  $x \in [x^C, x^W]$ ) with the same cost (in terms of mutations) as the  $\text{mon}(x^W, K)$ -tree in the proof of Theorem 1. We will see that, for this purpose, it is enough to modify the mentioned  $\text{mon}(x^W, K)$ -tree, by deleting certain arrows and adding transitions from higher to lower quantities. This will show stochastic stability of the whole interval. The first Lemma, though, establishes that this cannot be done with “direct” transitions.

**Lemma B.1.** *Let  $x \in [x^C, x^W]$ . No transition from  $\text{mon}(x, K)$  to  $\text{mon}(x', K)$  with  $x' < x$  is possible with two or fewer mutations.*

*Proof.* Transitions with one mutation are precluded by Lemma A.6. It suffices, hence, to consider two simultaneous mutations to  $x', x''$  with  $x' < x$ .

By Lemma A.7, the profit function  $g_p(z) = p \cdot z - C(z)$  is strictly concave and its unique maximum  $z(p)$  is increasing in  $p$ , with  $z(P(N \cdot x^W)) = x^W$ . Let  $Q = (N - 2)x + x' + x''$  be the total quantity produced after mutation, and let  $p = P(Q)$  be the corresponding market price.

If  $Q < Nx^W$ , then  $p > P(Nx^W)$  and hence  $z(p) \geq x^W$ . It follows that  $g_p(x') < g_p(x)$ , i.e. the  $x'$ -mutant earns less than the non-mutants *after* mutation (remember that  $N \geq 3$ ), and hence can not be successful.

If  $Q > Nx^W$ , then  $p < P(Nx^W) < P(Nx)$  and  $p \cdot x' - C(x') < P(Nx)x' - C(x')$ . Since  $z(P(Nx)) \geq x^W$ , it follows that  $P(Nx)x' - C(x') < P(Nx)x - C(x)$ . Combining both facts, we get that  $p \cdot x' - C(x') < P(Nx)x - C(x)$ , i.e. the  $x'$ -mutant earns less than the non-mutants *before* mutation and hence can not be successful. ■

Since transitions to lower quantities can not be achieved directly with two mutations, we must take a detour. The basic idea is that quantities in  $[x^C, x^W]$  can be first destabilized “upwards,” to high quantities above  $x^W$  from where we can add further transitions reaching lower quantities than the original ones. The next two lemmata concentrate on the first step in such transitions, giving us a range of higher quantities that we can reach from a given one (actually, from the corresponding monomorphic state).

**Lemma B.2.** *Let  $x \in [x^C, x^W]$ .<sup>13</sup> A transition from  $\text{mon}(x, K)$  to  $\text{mon}(y, K)$ , with  $x < y \leq 2 \cdot x$  can be achieved with two mutations if  $H(x, y) \geq 0$ , where*

$$H(x, y) = P(y + 0 + (N - 2) \cdot x)(y - x) - C(y) + C(x).$$

*Proof.* Consider the situation  $\text{mon}(x, K)$ , where all firms are producing the quantity  $x$ . Consider two simultaneous mutations to quantities  $x_1 = y, x_2 = 0$ , with  $x < y \leq 2 \cdot x$ . Call then  $Q = y + (N - 2) \cdot x \leq N \cdot x$ .

Note that  $P(Q)x - C(x) \geq P(N \cdot x)x - C(x)$ . It follows that the considered mutation can be successful if and only if  $P(Q)y - C(y) \geq P(Q)x - C(x)$ , i.e. if and only if  $H(x, y)$  is positive.<sup>14</sup> ■

<sup>13</sup>This Lemma holds actually for a larger range of quantities.

<sup>14</sup>The considered mutations (involving a mutant to quantity zero) might seem arbitrary.

Fix  $x$  and denote  $H_x(y) = H(x, y)$ . Note that  $H_x(x) = 0$  and that  $H'_x(x) = P((N-1) \cdot x) - C'(x) > P(N \cdot x) - C'(x) \geq 0$  (since  $x \leq x^W$ ). This proves that for  $y > x$  close enough to  $x$ , such a transition is possible. The problem is whether the transition is also possible for  $y > x$  “far away enough” for our purposes. The next lemma identifies the largest such  $y$  and shows several useful properties.

**Lemma B.3.** *For  $x \in [x^C, x^W]$ , define*

$$h(x) = \max\{y \in [x, 2 \cdot x] \mid H(x, y) \geq 0\}$$

*Then,  $h$  is continuous, and either  $h(x) = 2x$ , or  $h(x) \in (x, 2 \cdot x)$ . In the latter case,  $h(x)$  is implicitly defined by  $H(x, h(x)) = 0$ , and it is strictly decreasing.<sup>15</sup>*

The next lemma paves the way for the second step in the transitions explained above. Once we have reached a monomorphic state with a certain high quantity, we need to determine how far “down” we can go back. We are only interested in reaching quantities in  $[x^C, x^W]$ . Intuitively, we can concentrate on the relative advantages of mutants because they adjust their quantity in the direction of  $x^C$  and are hence better off in absolute terms. More rigorously, Lemma A.1 implies that a mutation from a higher to a lower quantity is always beneficial in absolute terms when it is so in relative terms. Therefore, we concentrate on the function  $D(x, \cdot)$ .

**Lemma B.4.** *For  $x > x^W$ , define*

$$f(x) = \min\{y \in [0, x] \mid D(x, y) \geq 0\}.$$

*Then,  $f$  is continuous, and either  $f(x) = 0$  or  $f(x) \in (0, x^W)$ . In the latter case,  $f$  is implicitly defined by  $D(x, f(x)) = 0$ , and it is strictly decreasing. Moreover,  $D(x, y) > 0$  for all  $y \in (f(x), x)$ .*

For high quantities (above  $x^W$ ), mutations to lower ones are beneficial in relative terms. The last lemma shows that, as we consider higher and higher quantities, the range of such beneficial deviations increases in size. In the next result we tackle a complementary consideration. Fixing a deviation  $y$ , we show that there exists a quantity  $\phi(y)$  such that the deviation  $y$  is beneficial for all monomorphic states corresponding to quantities above  $\phi(y)$ .

**Lemma B.5.** *The function  $D(x, y)$  is strictly convex in  $x$  for  $x > y$ . For all  $y < x^W$  there exists a unique  $\phi(y) > x^W$  such that  $D(\phi(y), y) = 0$  and  $D(x, y) > 0 \forall x > \phi(y)$ . Taking  $\phi(x^W) = x^W$ , the function  $\phi$  is continuous in  $(0, x^W]$ .<sup>16</sup>*

*In particular, if  $x_2 > x_1 > y$  and  $D(x_1, y) \geq 0$ , then  $D(x_2, y) > 0$ .*

Actually, a transition from  $\text{mon}(x, K)$  to  $\text{mon}(y, K)$  through two mutations to  $x_1 = y, x_2 \geq 0$  (such that  $x_1 > x, x_1 + x_2 \leq 2 \cdot x$ ) is possible if and only if the same transition can be achieved with two mutations to  $x_1$  and  $x'_2 = 0$ . To prove this, first note that if the transition is possible with the mutations to  $x_1, x_2$ , then  $P(x_1 + x_2 + (N-2) \cdot x)(x_1 - x) - C(x_1) + C(x) \geq 0$  which, since  $P$  is decreasing, implies that  $H(x, x_1) \geq 0$ . The reverse implication is trivial. It suffices to set  $x_2 = 0$ .

<sup>15</sup>Hence,  $h$  is either decreasing or “tent-shaped,” because  $2x$  is increasing and the implicit function defined by  $H(x, h(x)) = 0$  is decreasing.

<sup>16</sup>It is easy to see that the functions  $f$  and  $\phi$  are partial inverses, i.e.  $f(\phi(x)) = x$  for all  $x \in [0, x^W]$ , but  $\phi(f(x)) \neq x$  in general ( $f$  is not invertible).

Suppose that a monomorphic state  $\text{mon}(x, K)$  with  $x \in [x^C, x^W]$  has been destabilized with two mutations and a new state  $\text{mon}(x', K)$  has been reached. Lemma B.3 guarantees that this is possible for all  $x' \in (x, h(x))$ . The next lemma implies that, if  $x'$  is high enough, it must be the case that  $D(x', x)$  is strictly positive, and hence, from the new state  $\text{mon}(x', K)$ , a single mutation back to  $x$  or, by continuity, to an even lower quantity, will be beneficial in relative terms. This will be the core of the argument to show how to “go back” and construct transitions towards lower quantities in  $[x^C, x^W]$  (recall that we already know that  $\text{mon}(x^W, K)$  is stochastically stable).

**Lemma B.6.** *Let  $x \in [x^C, x^W]$ . Then,  $D(h(x), x) > 0$ . Moreover,  $h(x) > x^W$ .*

The next lemma (whose proof follows immediately from Lemmata B.3, B.5, and B.6) implies that, given any quantity  $x \in [x^C, x^W]$ , there exists a range of higher quantities  $y$  such that the corresponding monomorphic states can be reached from  $\text{mon}(x, K)$  with two mutations as in Lemma B.2, but new monomorphic states  $\text{mon}(x', K)$  with  $x' < x$  can be reached from  $\text{mon}(y, K)$  with a single mutation (see Figure 5).

**Lemma B.7.** *For all  $x \in [x^C, x^W]$ ,  $\phi(x) < h(x)$ .*

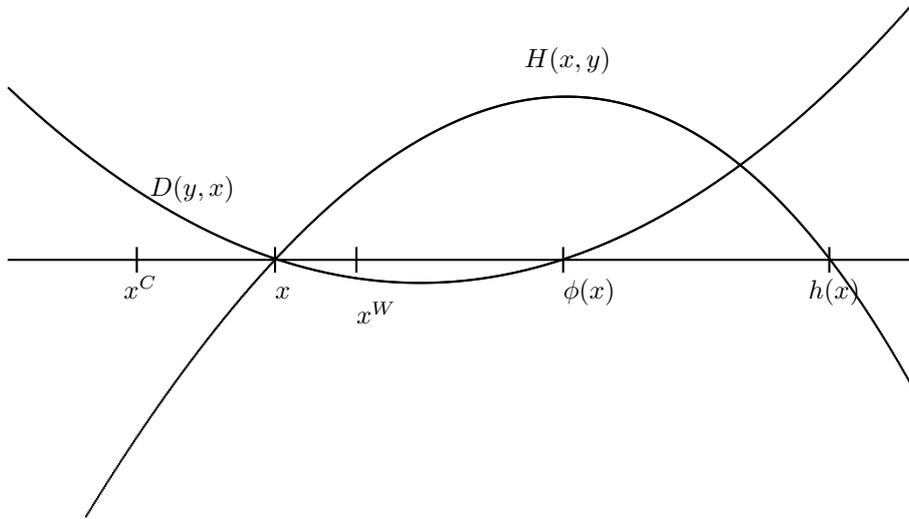


Figure 5: Two-step transitions. In a first step, a transition from  $\text{mon}(x, K)$  to  $\text{mon}(y, K)$ ,  $y > x$  can be achieved with two mutations if  $H(x, y) \geq 0$ . In a second step, a transition from  $\text{mon}(y, K)$  to  $\text{mon}(x, K)$  and to  $\text{mon}(x', K)$  with  $x' < x$  can be achieved with one mutation if  $D(y, x) > 0$ . The horizontal axis is the first variable of  $D(\cdot, \cdot)$  and the second of  $H(\cdot, \cdot)$ . The fact that  $\phi(x) < h(x)$  makes possible a net “downward” transition.

Now we are ready to prove the main theorem. The remaining proof shows that, for any  $x \in [x^C, x^W] \cap \Gamma$ , we can construct a  $\text{mon}(x, K)$ -tree of the same cost (i.e. number of involved mutations) as that of the  $\text{mon}(x^W, K)$ -tree constructed in the proof of Theorem 1. The proof is complicated by considerations regarding the discretization of the strategy space.

*Proof of Theorem 2.* Consider the  $\text{mon}(x^W, K)$ -tree of minimal cost constructed in the proof of Theorem 1. Theorem 2 will be proven if we can show that we can modify this tree into a  $\text{mon}(x, K)$  of the same cost for all  $x \in [x^C, x^W] \cap \Gamma$ . We distinguish two cases.

**Case 1:** For all  $x \in [x^C, x^W]$ ,  $h(x) \in (x, 2x)$ .

In this case,  $h(x) > x$  is a strictly decreasing function, implicitly defined by  $H(x, h(x)) = 0$  (see Lemma B.3). Moreover,  $h(x) \geq h(x^W) > x^W$  for all  $x \in [x^C, x^W]$ .

Let  $\delta_h = \min\{h(x) - \phi(x) \mid x \in [x^C, x^W]\}$ . Since  $h$  and  $\phi$  are continuous by Lemmata B.3 and B.5 and  $h(x) > \phi(x)$  for all  $x \in [x^C, x^W]$  by Lemma B.7, we know that  $\delta_h > 0$ .

We can define a continuous function  $\psi$  by

$$\psi(x) = \frac{1}{2}(h(x) + \phi(x)) \in (\phi(x), h(x))$$

Let  $x \in [x^C, x^W]$ . By Lemma B.5,  $\phi(x) > x^W > x$  with  $D(\phi(x), x) = 0$ . By Lemma B.4, since  $D(\phi(x), x) = 0$  it must be that  $f(\phi(x)) = x$  and  $f$  is strictly decreasing at  $\phi(x)$ . Since  $\phi(x) < \psi(x) < h(x)$  and  $f(\phi(x)) = x > 0$ , it follows from Lemma B.4 that  $x = f(\phi(x)) > f(\psi(x)) \geq f(h(x))$ . For  $x = x^W$ ,  $\phi(x^W) = x^W$  and hence  $f(\phi(x^W)) = x^W > f(\psi(x^W))$ . In summary,  $f(\psi(x)) < x$  for all  $x \in [x^C, x^W]$ .

Let  $\delta_f = \min\{x - f(\psi(x)) \mid x \in [x^C, x^W]\}$ . Since  $f$  and  $\psi$  are continuous by Lemmata B.3, B.4, and B.5 and  $x > f(\psi(x))$  for all  $x \in [x^C, x^W]$ , we know that  $\delta_f > 0$ .

Let  $0 < \delta < \frac{1}{3} \min\{\delta_h, \delta_f\}$ . Consider  $x_0 = x^W$ . Let  $y_0 \in (\psi(x_0), h(x_0)) \cap \Gamma$  (which exists because  $\delta < \frac{1}{3}\delta_h$ ). Since  $H_x(y) = H(x, y)$  is strictly concave in  $y$  by Lemma B.3, we know that  $H(x_0, y_0) > 0$ , i.e. a transition from  $\text{mon}(x_0, K)$  to  $\text{mon}(y_0, K)$  with two mutations is possible.

Let  $x_1 \in (f(\psi(x_0)), x_0) \cap \Gamma$ . It exists because  $\delta < \delta_f$ . Since  $f$  is decreasing,  $f(y_0) \leq f(\psi(x_0)) \leq x_1$ , which implies that  $D(y_0, x_1) \geq 0$ , i.e. the transition from  $\text{mon}(y_0, K)$  to  $\text{mon}(x_1, K)$  is possible with one mutation.

Consider the  $\text{mon}(x^W, K)$ -tree of minimal cost constructed in the proof of Theorem 1. Delete the arrows leaving  $\text{mon}(x_1, K)$  (which has cost 2) and  $\text{mon}(y_0, K)$  (cost 1). Replace them with the arrow from  $\text{mon}(x_0, K)$  to  $\text{mon}(y_0, K)$  (cost 2) and the arrow from  $\text{mon}(y_0, K)$  to  $\text{mon}(x_1, K)$  (cost 1). This yields a  $\text{mon}(x_1, K)$ -tree with the same cost as the original one and hence proves that  $\text{mon}(x_1, K)$  is stochastically stable. Moreover, by Lemma A.8 all states  $\text{mon}(x, K)$  with  $x$  in  $[x_1, x^W] \cap \Gamma$  are also stochastically stable.

Notice that the grid is a priori fixed and  $x_1 < x_0$ , i.e.  $x_1 \leq x_0 - \delta$ . This procedure has added at least one stochastically stable state. Repeating the argument from  $x_1$ , we obtain a new stochastically stable state  $\text{mon}(x_2, K)$  with  $x_2 \leq x_1 - \delta$ . After a finite number of iterations, we obtain that  $\text{mon}(x^C, K)$  is stochastically stable, and hence all monomorphic states with quantities in  $[x^C, x^W]$  are stochastically stable.

**Case 2:** For some  $x \in [x^C, x^W]$ ,  $h(x) = 2x$ .

Apply the same procedure as in Case 1. Let  $x_i$  be the first quantity in the sequence constructed there such that  $h(x_i) = 2x_i$ . It follows that  $h(x) = 2x > x^W$  for all  $x \in [x^C, x_i]$  (because  $h$  is continuous by Lemma B.3, strictly decreasing whenever  $h(x) \in (x, 2x)$ , and  $2x$  is strictly increasing).

By Lemma B.4  $f$  is decreasing. Since  $2x_i \geq 2x^C$ , it follows that  $f(2x_i) \leq f(2x^C)$ . Since  $h(x^C) = 2x^C$  and  $D(h(x^C), x^C) > 0$  by Lemma B.6, we know that  $f(2x^C) < x^C$ . Hence,  $f(2x_i) < x^C$  and  $D(2x_i, x^C) > 0$ .

We can set now  $x_{i+1} = x^C$  and complete the previous tree with an arrow from  $\text{mon}(x_i, K)$  to  $\text{mon}(2x_i, K)$  (at cost 2) and an arrow from  $\text{mon}(2x_i, K)$  to  $\text{mon}(x^C, K)$  (at cost 1). ■

*Acknowledgements.* This paper has benefitted from helpful comments by Ana B. Ania, Georg Kirchsteiger, Manfred Nermuth, Jörgen Weibull, two anonymous referees, the editor, Simon Anderson, and seminar participants at Universitat Autònoma de Barcelona, Universidad Carlos III de Madrid, and the University of Vienna. Financial support from the European Science Foundation and the Austrian FWF (under Project P15281) is gratefully acknowledged.

## References

- Alchian, A., 1950. Uncertainty, Evolution, and Economic Theory. *Journal of Political Economy* 57, 211-221.
- Alós-Ferrer, C., 2000. Learning, Memory, and Inertia. Working Paper 0003, Department of Economics, University of Vienna.
- Alós-Ferrer, C., Ania, A.B., Vega-Redondo, F., 1999. An Evolutionary Model of Market Structure. In: Herings, P.J.J., Talman, A.J.J., van der Laan, G. (Eds), *Theory of Markets*. North-Holland, Amsterdam, pp. 139-163.
- Ania, A.B., Tröger, T., Wambach, A., 2002. An Evolutionary Analysis of Insurance Markets with Adverse Selection. *Games and Economic Behavior*, 40, 153-184.
- Alós-Ferrer, C., Ania, A.B., Schenk-Hoppé, K.R., 2000. An Evolutionary Model of Bertrand Oligopoly. *Games and Economic Behavior*, 33, 1-19.
- Barron, G., Erev, I., 2003. Small Feedback-Based Decisions and their Limited Correspondence to Description-Based Decisions. *Journal of Behavioral Decision Making*, forthcoming.
- Bergin, J., Bernhardt, D., 1999. Comparative Dynamics. Institute for Economic Research Working Paper 981, Queen's University at Kingston (Canada).
- Conlisk, J., 1980. Costly Optimizers versus Cheap Imitators. *Journal of Economic Behavior and Organization*, 275-293.
- Crawford, V., 1991. An 'Evolutionary' Interpretation of Van Huyck, Battalio, and Beil's Experimental Results on Coordination. *Games and Economic Behavior*, 3, 25-59.
- Delgado, J., Moreno, D., 2002. Coalition-proof Supply Function Equilibria in Oligopoly. Mimeo, Department of Economics, Universidad Carlos III de Madrid.
- Ellison, G., 2000. Basins of Attraction, Long Run Stochastic Stability, and the Speed of Step-by-Step Evolution. *Review of Economic Studies*, 67, 17-45.

- Erev, I., Barron, G., 2003. On Adaptation, Maximization, and Reinforcement Learning among Cognitive Strategies, Technion Working Paper.
- Erev, I., Roth, A.E., 1998. Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria. *American Economic Review*, 88, 848-81.
- Ferreira, J.L., 2003. Strategic interaction between futures and spot markets. *Journal of Economic Theory*, 108, 141-151.
- Freidlin, M., Wentzell, A., 1988. *Random Perturbations of Dynamical Systems*, 2nd Edition. Springer Verlag, New York.
- Fudenberg, D., Levine, D., 1998. *The Theory of Learning in Games*. MIT Press, Cambridge, MA.
- Gale, D., and Rosenthal, R.W., 1999. Experimentation, Imitation, and Stochastic Stability. *Journal of Economic Theory*, 84, 1-40.
- Hehenkamp, B., 2002. Sluggish Consumers: An Evolutionary Solution to the Bertrand Paradox. *Games and economic Behavior*, 40, 44-76.
- Kaelbling, L.P., Littman, M.L., Moore, A.P., 1996. Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, 4, 237-285.
- Kandori, M., Mailath, G., Rob, R., 1993. Learning, Mutation, and Long-Run Equilibria in Games, *Econometrica*, 61, 29-56.
- Kandori, M., Rob, R., 1995. Evolution of Equilibria in the Long Run: a general theory and applications. *Journal of Economic Theory*, 65, 383-414.
- Nöldeke, G., Samuelson, L., 1997. A Dynamic Model of Equilibrium Selection in Signaling Markets. *Journal of Economic Theory*, 73, 118-156.
- Oechssler, J., 2002. Cooperation as a Result of Learning with Aspiration Levels. *Journal of Economic Behavior and Organization*, 49, 405-409.
- Pingle, M., Day, R.H., 1996. Modes of economizing behavior: Experimental evidence. *Journal of Economic Behavior and Organization*, 29, 191-209.
- Rhode, P., Stegeman, M., 2001. Non-Nash Equilibria of Darwinian Dynamics with Applications to Duopoly. *International Journal of Industrial Organization*, 19, 415-453.
- Samuelson, L., 1997. *Evolutionary games and equilibrium selection*. MIT Press, Cambridge, MA.
- Sarin, R., 2000. Decision Rules with Bounded Memory. *Journal of Economic Theory*, 90, 151-160.
- Schaffer, M.E., 1988. Evolutionarily Stable strategies for a Finite Population and a Variable Contest Size. *Journal of Theoretical Biology*, 132, 469-478.
- Schaffer, M.E., 1989. Are Profit-Maximisers the Best Survivors? *Journal of Economic Behavior and Organization*, 12, 29-45.

- Schenk-Hoppé, K.R., 2000. The Evolution of Walrasian Behavior in Oligopolies. *Journal of Mathematical Economics*, 33, 35-55.
- Schipper, B., 2001. Imitators and Optimizers in Symmetric n-Firm Cournot Oligopoly. Mimeo, University of Bonn.
- Sutton, R.S., Barto, A.G., 1998. Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA.
- Tanaka, Y., 1999. Long Run Equilibria in an Asymmetric Oligopoly. *Economic Theory*, 14, 705-715.
- Tanaka, Y., 2000. Stochastically stable states in an oligopoly with differentiated goods: equivalence of price and quantity strategies. *Journal of Mathematical Economics*, 34, 235-253.
- Tanaka, Y., 2001. Evolution to equilibrium in an asymmetric oligopoly with differentiated goods. *International Journal of Industrial Organization*, 19, 1423-1440.
- Thijssen, J., 2001. Stochastic Stability of Cooperative and Competitive Behavior in Cournot Oligopoly. Discussion Paper 2001-19, CentER, Tilburg.
- Vega-Redondo, F., 1997. The Evolution of Walrasian Behavior, *Econometrica*, 65, 375-384.
- Williams, G.A., Miller, R.B., 2002. Change the Way you Persuade. *Harvard Business Review*, 80 (5), 64-73.
- Young, P., 1993. The Evolution of Conventions, *Econometrica*, 61, 57-84.
- Young, P., 1998. Individual Strategy and Social Structure. Princeton University Press, Princeton, NJ.