

Evolving Rules: Imitation and Best Response Learning in Cournot Oligopoly ^{*}

Xiaomeng Ding[†] Simon Weidenholzer[‡] Boyu Zhang[§]

November 14, 2025

Abstract

We study evolutionary dynamics in which firms endogenously revise the behavioral rules that govern strategy revisions in symmetric Cournot oligopoly. Specifically, we consider two principles that guide rule revision, *No-Birth* and *Survival-of-the-Fittest*, both grounded in imitation-based heuristics. We show that, under these principles, all firms eventually adopt the same behavioral rule. Focusing on two classical rules, myopic best response and imitation, we demonstrate that rule revision plays a crucial role in determining long-run equilibria in Cournot oligopoly. The set of long-run equilibria includes the state where all players use best response learning and choose the Nash equilibrium quantities and states where all firms use imitation learning and choose specific symmetric quantities which include (but are not necessarily restricted to) Walrasian quantities. Our results extend to more general aggregative games.

Keywords: Behavioral Rules, Rules Evolution, Long-run Equilibrium, Cournot oligopoly

^{*}We are indebted to Dan Friedman and Jörg Oechssler for helpful comments. This work was conducted while Xiaomeng Ding was visiting the Department of Economics at University of Essex, whose hospitality is greatly acknowledged. Ding acknowledges financial support of the China Scholarship Council program (No. 202406040148). Zhang acknowledges financial support from the National Natural Science Foundation of China (No. 72131003), the Beijing Natural Science Foundation (No. Z220001), and the National Key Research and Development Program of China (No. 2020YFA0712900). All authors contributed equally.

[†]Laboratory of Mathematics and Complex Systems, Ministry of Education, School of Mathematical Sciences, Beijing Normal University, 100875 Beijing, P. R. China. Email: dingxm0616@gmail.com.

[‡]Department of Economics, University of Essex, CO4 3SQ, Colchester, UK.

[§]Laboratory of Mathematics and Complex Systems, Ministry of Education, School of Mathematical Sciences, Beijing Normal University, 100875 Beijing, P. R. China. Email: zhangby@bnu.edu.cn.

1 Introduction

Often individuals use simple behavioural rules to determine their actions in strategic settings (Nax et al., 2016; Huck et al., 2017; Kovářik et al., 2018; Newton, 2018). But different behavioral rules may prescribe different actions and subsequently lead to different societal outcomes. We are consequently interested to understand which rules economic agents come to adopt and what the implications are for action choice when agents choose rules that have historically performed well for themselves and others.¹

We cast our research question in the context of Cournot competition which is the workhorse model of imperfect competition in Industrial Organization but may also more widely be considered as a metaphor for social dilemmas with a conflict between maximizing individual outcomes and achieving cooperative outcomes. In this setting firms can either choose their actions via imitation (e.g. Vega-Redondo, 1997) or through myopic best response (e.g. Fisher, 1961). Under imitation learning firms simply choose the quantity that has yielded the highest profit in the industry in the previous period, and under best response learning, they play a best response to the previous distribution of quantities. When firms consider switching rules, we are interested in scenarios where they are guided by two simple principles: the *No-Birth* principle, which postulates that no new rule may emerge other than those already in use in the industry and the *Survival-of-the-Fittest* principle, which assumes that rules that have earned the highest average historical payoffs to some firm are adopted with positive probability. These two principles capture a broad class of criteria grounded in imitation-based heuristics, where rules with the highest historical payoffs are adopted. Rule adoption, however, is not necessarily confined to these top rules; for example, a rule that outperforms a firm’s current choice may also be selected. We augment this adjustment process by allowing for mistakes in action choice and in rule choice. Considering the case when the probability of mistakes vanishes allows us to characterize the set of *long-run equilibria* (Kandori et al., 1993; Young, 1993) as predictions for the long-run behaviour of the dynamics. Following this literature, the set of long-run equilibria can be identified by studying the relative robustness of absorbing

¹Rule revision captures more widespread phenomena than decision makers occasionally adjusting their behavioral rules. In business contexts, firms or institutional agents often operate with finite managerial tenure and routine leadership turnover, circumstances that naturally create opportunities for reassessing past performance and adjusting decision-making rules. See e.g., Denis and Denis (1995) and Weisbach (1995) for empirical studies documenting changes in firms’ strategies following managerial turnover.

sets to mistakes.

To appreciate and understand our results, it is useful to consider the benchmark cases where i) all firms use myopic best response to determine their quantities and ii) all firms use imitation learning. In the first case, [Cournot \(1836\)](#) has already discussed convergence of a myopic best response process to the Cournot Nash equilibrium under linear demand and costs.² In the second case where firms imitate quantities that have yielded the highest profit in the previous period, [Vega-Redondo \(1997\)](#) has shown that the Walrasian equilibrium, where firms act as (if they were) price takers, is the unique long-run equilibrium. To understand this, note that the Walrasian quantity has the property that any firm choosing it will enjoy a payoff advantage over firms choosing different quantities. Thus, any mistake towards the Walrasian quantity will be imitated by other firms whereas a single mistake away from the profile where all firms choose the Walrasian quantity will not be followed. Note that the Walrasian equilibrium -underpinned by imitation- is more competitive than the Cournot Nash equilibrium arising under myopic best response learning; the former yields strictly higher output and lower profit than the latter.

In the case where firms may adjust their rules, it is useful to observe that under the *No-Birth-* and the *Survival-of-the-Fittest-* principle the dynamics will converge to states where all firms use the same behavioral rule. If the behavioral rule used by all firms is best response learning, then in fact all firms will have to choose the Nash quantity. In contrast, when all firms use imitation learning, states in which all firms choose the same quantity are absorbing.

Given this characterization of absorbing sets, we then proceed to identify the number of mistakes required for transitions among these states. At the core of our results is the fact that only one mistake in rule choice is required to move from the state where all firms are imitators and choose the Walrasian quantity to the state where all firms are best responders and choose the Nash quantity. At the same time, it is possible to move from the state where all firms are best responders choosing the Nash quantity to the state where they are all imitators choosing the Nash quantity. Starting from the latter state, we can connect a whole array of states where firms are imitators via different mistakes in action choice. While these action mistakes may be directly towards the Walrasian quantity, they

²Subsequent studies show that myopic best response dynamics also converge to the Nash equilibrium for more general demand and cost functions ([Dubey et al., 2006](#); [Vives, 1999](#), Chapter 4). In addition, the convergence result also holds for other belief-based learning rules, e.g., the fictitious play ([Fudenberg and Levine, 1998](#), Chapter 2), the Bayesian learning ([Kalai and Lehrer, 1993](#)).

can also be towards other quantities that are likewise relative-payoff improving. Putting these observations together allows us to characterize the set of long-run equilibria. If mistakes in rule choice happen less frequently than in action choice, then we have two types of long-run equilibria: in one, everybody is a best responder and chooses the Nash quantity, and in the other, everybody is an imitator choosing the Walrasian quantity. If, however, action and rule mistakes occur at the same rate, then in addition to the two types above, states where all firms are imitators all choosing the same quantity from an interval are long-run equilibria.³

These results are surprising, especially given that imitators outperform best responders in the long run when the two types of learners coexist (Schipper, 2009).^{4,5} The intuition is that imitators, acting as price takers, choose a quantity that yields the highest relative payoff, while best responders behave like Stackelberg followers. Consequently, it is natural to conjecture that, imitation learning drives out best response learning under payoff monotone selection dynamics. Our analysis, however, reveals that best responders can also dominate the population in the long run. The key mechanism is that the average historical payoff of imitators choosing the Walrasian equilibrium may (at least temporarily) be dominated by the payoff received by best responders.

Another implication of our results is that in the presence of rule revisions the Walrasian equilibrium is not necessarily the unique long-run outcome when all firms employ imitation learning. Specifically, when action and rule mistakes occur at the same frequency, the Cournot Nash equilibrium (arising from best response learning) serves as an intermediate stability plateau, facilitating deviation from the Walrasian equilibrium and giving rise to a spectrum of output levels that may even include the collusive quantity. This contrasts with the well-established results based on imitation which typically hold learning rules fixed, and highlights the role of endogenous rule revision in shaping long-run outcomes.

³If there are only two firms, the lower bound of this interval is the Nash quantity and the upper bound is a specific quantity above the Walrasian quantity. If there are more than two firms and the grid of feasible output is sufficiently fine, the lower bound is zero while the upper bound is a specific quantity strictly greater than the Walrasian quantity.

⁴Schipper (2009) shows that any Pseudo-Stackelberg state where imitators are weakly better off than best responders is absorbing. Moreover, when mistakes are possible, the Nash equilibrium cannot be the unique prediction and hence the imitators' long run average per-period payoff is strictly higher than that of best responders.

⁵Further support for the superior performance of imitation learning over other rules is provided by Hehenkamp and Kaarbøe (2008), Duersch et al. (2010, 2012) or Bossan et al. (2015).

Related literature

Early theoretical work on learning or evolutionary approaches to Cournot competition assumes that all firms adhere to the same behavioral rule, such as myopic best response or imitation learning, with long-run outcomes depending on the rule: the former converges to the Nash equilibrium, whereas the latter yields the Walrasian outcome. However, experimental studies indicate that the coexistence of multiple rules within the same environment, rather than the dominance of a single behavioral rule (Huck et al., 1999, 2002; Bosch-Domènech and Vriend, 2003; Apesteguia et al., 2010; Bigoni and Fort, 2013). In particular, the coexistence of both best response and imitation learning is frequently observed, with the prevalence of each affected by the information settings. Furthermore, recent evidence from long-horizon experiments suggests that subjects adjust their behavioral rules over time (Friedman et al., 2015; Oechssler et al., 2016), even though the rule adjustment is significantly affected by information feedback (Huck et al., 2017). In a related vein, Alós-Ferrer and Ritschel (2021) provide evidence that, in Cournot oligopolies, imitation and more deliberative rules such as myopic best response coexist at the intra-individual level, with the rules used determined by cognitive interaction.

Departing from the assumption of a single behavioral rule, Schipper (2009) studies a theoretical model in which a fixed set of best responders and a fixed set of imitators repeatedly play a symmetric Cournot oligopoly. The analysis shows that the long-run distribution converges to a subset of those states corresponding to Stackelberg equilibria, where imitators earn a strictly higher average per-period payoff than best responders. Building on this finding, Schipper (2009) conjectures that if agents were allowed to revise their rules according to past performance, and if the occurrence of rule revision were sufficiently infrequent relative to quantity revision, the share of imitators would increase and eventually dominate the population. To the best of our knowledge, however, no study has yet provided an explicit theoretical prediction of how the two rules evolve in the context of Cournot oligopolies.⁶ In our study, we explicitly define a rule revision protocol that refers to historical average payoffs and provide a complete characterization of the long-run outcomes. Our analysis, however, challenges the conjecture put forward by Schipper (2009) that only imitators can be around in the long run.

⁶Thijssen (2009) is closely related to our work in terms of motivation. However, the paper only establishes approximate results under the assumption of an extremely low frequency of rule revision and there is no explicit definition of the rule revision procedure.

Juang (2002) examines rule evolution in a setting where agents are randomly matched into pairs to play 2×2 coordination games, with some adopting best response learning and others relying on imitation learning.⁷ When behavioral rules are fixed, the analysis shows that the long-run outcome is determined by the relative proportion of the two rules in the population. In contrast, when agents are allowed to revise their rules according to the *experimental* criterion, they select with positive probability among the rules currently in use in the population without reference to historical performance, and in the long run all agents coordinate at the Pareto-efficient equilibrium, with the prevailing rule being either imitation or best response. We advance Juang (2002)’s framework by explicitly conceptualizing the incentives underlying rule revision.

A more recent contribution by Juang and Sabourian (2021) introduces the *monotonicity* principle, under which agents take the performance of rules into account. While sharing some similarities, our study differs in several important dimensions. First, we restrict attention to symmetric Cournot oligopolies and focus on two classical behavioral rules, best response and imitation, motivated by robust experimental evidence. By contrast, Juang and Sabourian (2021) consider a random matching framework in which agents, in every period, are randomly matched into groups and assigned random roles to play normal-form games. Besides, a behavioral rule is defined as a mapping from a finite set of histories to the action set so that the number of possible rules grows exponentially with memory length. Second, under our *Survival-of-the-Fittest* principle, each agent evaluates the performance of its current rule relative to others using past average payoffs. In contrast, under their *monotonicity* principle, agents make period-by-period comparisons of their own payoff with that of agents playing the same role across different matches, and only a rule that has outperformed any alternative in every period and match is regarded as attractive. Third, Juang and Sabourian (2021) assume that mistakes happen only during rule revision. Moreover, these mistakes are payoff-dependent, which is crucial for their main result that only action profiles that maximize the minimum payoff among agents are selected in the long run when the set of feasible rules is sufficiently rich.⁸

⁷Note that Juang (2002)’s definition of imitators differs slightly from ours: in his model, imitators select the action that yields the highest average payoff over the population.

⁸For example, in the Cournot game studied in this paper, the unique action profile that maximizes the minimum payoff among agents is given by the profile where all firms produce the collusive quantity. It is worth noting, however, that the results of Juang and Sabourian (2021) do not apply for Cournot games when the set of feasible rules is restricted to best response and imitation learning (as studied in the present contribution).

Some studies are conceptually related to ours but methodologically different. [Conlisk \(1980\)](#) develops a discrete-time population dynamics with two behavioral types. One type pays a fixed cost to adapt directly to the stochastically evolving social state, while the other type incurs no cost but can only follow a lagged social convention, thereby suffering losses from its deviation relative to the true state. In this model, the performance of behavioral rules is assessed by the average per-agent payoffs of the two types in the previous period, which monotonically determine their respective population shares. [Droste et al. \(2002\)](#) considers an infinite population of firms randomly matched each period to play Cournot duopolies. Firms choose from a finite set of behavioral rules in each period, with population shares evolving according to a replicator equation with mutation noise, where realized profits from the previous period serve as the measure of fitness. Their results show that the coevolution of population and quantity dynamics in Cournot oligopolies can be complex and endogenous fluctuations may emerge in the long run. Beyond methodological differences, the main distinction from our study is that, in these studies, rule shares are updated each period, leaving no scope for assessing rule performance over longer horizons.

The rest of this paper is organized as follows. Section 2 introduces the model and the dynamics of action and rule revision. Section 3 conducts a long-run analysis and presents the results. Section 4 concludes. All proofs are relegated to Appendix A. Appendix B pursues the extension to aggregative games.

2 The model

2.1 The base game

We are primarily interested in symmetric Cournot games where n firms produce a homogeneous good.⁹ Denote by $N = \{1, \dots, n\}$ the set of firms and assume $n \geq 2$. Letting q_i denote the output of firm i and $Q := \sum_{i \in N} q_i$ denote the total industry output, the profit of firm i is given by $\pi(q_i, Q) = p(Q)q_i - c(q_i)$ where $p : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ represents the inverse demand function and $c : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ the cost function.

To facilitate our analysis, we impose standard assumptions common in Cournot oligopoly (see, e.g., [Tirole, 1988](#), Chapter 5). Specifically, the inverse demand function p is assumed to be twice differentiable on the closed interval $[0, Q_{\max}]$, where $p(0) > 0$ and $p(Q) = 0$

⁹Appendix B generalizes our analysis to aggregative games.

for all $Q \geq Q_{\max}$. Furthermore, p is strictly decreasing, $p'(\cdot) < 0$, and weakly concave, $p''(\cdot) \leq 0$ on $[0, Q_{\max}]$. Moreover, the cost function c is assumed to be twice differentiable on $\mathbb{R}_+$, strictly increasing, $c'(\cdot) > 0$, and weakly convex, $c''(\cdot) \geq 0$. To ensure a nontrivial market, we assume $c'(0) < p(0)$. The best response correspondence of firm i gives the set of profit maximizing quantities when the combined output of the other firms is given by Q_{-i} and is denoted by

$$b_i(Q_{-i}) = \arg \max_{q \in \mathbb{R}_+} \pi(q, q + Q_{-i}).$$

Note that under our assumptions, the profit function is single peaked so that b_i is a function.

The following two quantities will be of special interest.

Definition 1. *The Cournot Nash quantity q^N satisfies $p(nq^N)q^N - c(q^N) \geq p((n-1)q^N + q)q - c(q)$ for all $q \in \mathbb{R}_+$.*

Definition 2. *The Walrasian quantity q^W satisfies $p(nq^W)q^W - c(q^W) \geq p(nq^W)q - c(q)$ for all $q \in \mathbb{R}_+$.*

Under the above assumptions, the existence and uniqueness of a strictly positive Cournot Nash quantity is well established (Tirole, 1988). The Walrasian quantity is also strictly positive and unique (Vega-Redondo, 1997). Moreover we have $q^W > q^N > 0$. We denote the profits made when everybody chooses the Cournot Nash output by $\pi^C = \pi(q^C, nq^C)$ and the profits when everybody chooses the Walrasian output by $\pi^W = \pi(q^W, nq^W)$.

For technical reasons, we assume that each firm selects quantity from a common finite strategy set $\Gamma = \{0, \lambda, 2\lambda, \dots, \nu\lambda\}$ as its output, where $\lambda > 0$ is small and $\nu \in \mathbb{N}$ is large. We further impose that this grid includes q^N and q^W .

2.2 The dynamic process

We are interested in a dynamic process where firms periodically i) use simple behavioral rules to determine their output in the Cournot oligopoly and ii) decide which rule to use. We proceed to outline the details of this process. Figure 1 provides a sketch.

Time is discrete, $t = 1, 2, \dots$. At the end of each period t , first, with probability $\gamma \in (0, 1)$ firm i receives opportunity to revise its rule used for action choice. Such rule

revision opportunities are assumed to be independent of firms and time. When receiving a rule revision opportunity, firm i applies a rule selection criterion to decide on its future rule for action revision (specified below), denoted by $\rho(t+1)$. With probability $1-\gamma$ the firm does not receive the opportunity to revise its rule, in which case it sticks to its current rule. Note that the expected number of periods between two successive rule revisions for each firm is $1/\gamma$.

Following this, with probability $\theta \in (0,1)$, firm i receives opportunity (independent across firms, time and rule revision opportunities) to update its output. Upon receiving such an action revision opportunity, firm i follows its behavioral rule, $\rho_i(t+1)$, that prescribes an output for the next period. With the remaining probability $1-\theta$, firm i does not receive the opportunity to revise and retains its current output.

Our process thus exhibits *inertia* in rule revision and in quantity revision. We remark that inertia in quantity revision¹⁰ is required to rule out certain cyclical behaviors¹¹ and that our main results also hold under alternative assumptions on strategy revision opportunities.¹²

2.2.1 Action revision

When revising output, firms do so following behavioral rules. We denote the output profile of firms in period t by $\mathbf{q}(t) = (q_1(t), q_2(t), \dots, q_n(t)) \in \Gamma^n$, and define $Q(t) = \sum_{i \in N} q_i(t)$. The corresponding profit profile is $\boldsymbol{\pi}(t) = (\pi_1(t), \pi_2(t), \dots, \pi_n(t)) \in \mathbb{R}^n$, where $\pi_i(t) = \pi(q_i(t), Q(t))$ for all $i \in N$. Let $\mathcal{R}$ denote the set of behavioral rules available to firms and denote by $\boldsymbol{\rho}(t) = (\rho_1(t), \rho_2(t), \dots, \rho_n(t)) \in \mathcal{R}^n$ the profile of rules adopted by firms in period t . While our framework can accommodate more general settings, our analysis focuses on a scenario where firms determine their output through either (myopic) best response learning or through imitation. Letting BR denote best response learning and IM imitation learning, we have $\mathcal{R} = \{BR, IM\}$.

When a firm employs BR , it chooses a quantity that maximizes its profit given the

¹⁰See also [Huck et al. \(2002\)](#) documenting the presence of inertia in quantity revision in experimental Cournot markets.

¹¹See e.g., [Theocharis \(1960\)](#).

¹²For instance, one may assume that a firm uses a revised rule immediately to choose its output or that it always adjusts either its rule or its output, but not both in the same period.

output of other firms in the previous period.¹³ Formally, if $\rho_i(t+1) = BR$,

$$q_i(t+1) \in \arg \max_{q \in \Gamma} \pi_i(q, q + Q_{-i}(t)),$$

where $Q_{-i}(t) := \sum_{j \in N \setminus \{i\}} q_j(t)$ is the combined output of other firms in period t . If there are multiple quantities in this set, the firm is assumed to choose all of its best replies with positive probability.¹⁴

When a firm follows IM , it chooses a quantity that has earned the highest profit in the previous period among all firms in the industry.¹⁵ Formally, if $\rho_i(t+1) = IM$, then

$$q_i(t+1) = q_j(t) \quad \text{with} \quad j \in \arg \max_{l \in N} \pi_l(t).$$

As before, the firm puts positive probability in all quantities in this set.

2.2.2 Rule revision

When firms decide on which behavioral rule to use, each is guided by general principles that capture a set of criteria. In broad terms, firms look back at the historical performance of rules and then choose well-performing rules.

For each firm i , let $\tau_i(t) \in \mathbb{N}$ be the number of periods it has used its current behavioral rule (since its last rule revision) looking back from period t . For technical reasons we have to truncate the number of periods firms look back.¹⁶ To this end, assume that firms take at most M previous periods into account and note that M can be chosen arbitrarily large (but finite). Denote by $\tilde{\tau}_i(t) = \min\{\tau_i(t), M\}$ the number of periods that firm i effectively takes into consideration.

In particular, each firm i tracks the average performance of its current rule since its last rule revision by $f_i(t)$, which is defined as the average profit¹⁷ a firm obtains during

¹³See e.g. Ellison (1993, 2000), Blume (1993), and Oyama et al. (2015).

¹⁴Note that under our assumption the best response function is single peaked, so that in combination with our finite quantity grid there can be at most two best response quantities.

¹⁵This rule is known as the *Imitate-the-Best-Max* rule, see e.g. Vega-Redondo (1997), Alós-Ferrer and Weidenholzer (2006, 2008), or Apesteguia et al. (2010). We remark that most of our results would go through when considering a larger class of imitative rules such as in e.g. Apesteguia et al. (2007) and Alós-Ferrer and Weidenholzer (2014).

¹⁶Specifically, we have to ensure that the state space is finite.

¹⁷One may instead consider discounted average profits. In this case, let $\delta \in (0, 1]$ be the discount

the past $\tilde{\tau}(t)$ periods. Formally,

$$f_i(t) = \frac{\sum_{r=t-\tilde{\tau}_i(t)+1}^t \pi_i(r)}{\tilde{\tau}_i(t)}.$$

With this definition at hand, we define a general class of criteria applied for rule revision. Specifically, we are interested in criteria fulfilling the following two intuitive principles.

No-Birth (NB). A revising firm i may adopt a rule in period $t + 1$ only if that rule is already in use by at least one firm in the industry in period t . Formally, if $\rho_i(t + 1) = \rho$, $\rho \in \mathcal{R}$, then there exists $j \in N$ such that $\rho_j(t) = \rho$.

Survival-of-the-Fittest (SF). A revising firm i switches with positive probability to any of the rules employed by the firms attaining the highest average profits. Formally, $\rho_i(t + 1) = \rho_j(t)$ with positive probability if $j \in \arg \max_{l \in N} f_l(t)$.

The behavioral rationale underlying these two principles is inherently imitative and embodies the customary monotonicity considerations contemplated by evolutionary theory, while rules not adopted by any firm are eliminated from consideration and do not reemerge in future revisions.¹⁸

To illustrate our two principles, we now showcase several examples.

Example 1. *Under the Imitate-the-Best-Max criterion a revising firm employs a behavioral rule that has yielded the highest average profit in the industry. Formally, firm i*

factor and define

$$f_i(t) = \frac{\sum_{r=t-\tilde{\tau}_i(t)+1}^t \delta^{t-r} \pi_i(r)}{\sum_{r=t-\tilde{\tau}_i(t)+1}^t \delta^{t-r}}.$$

All results in the paper remain valid under this alternative assumption.

¹⁸The *No-Birth* principle corresponds to the invariance property of many deterministic evolutionary dynamics (e.g. replicator dynamics), whereby once a strategy dies out there is no way to revive it, and hence each face of the state space (a simplex) is invariant (Hofbauer and Sigmund, 1998). In finite populations, related properties are implied by Moran- or Pairwise-Comparison- processes (Wang et al., 2023). Juang and Sabourian (2021) also assume the *No-Birth* principle.

uses

$$\rho_i(t+1) = \rho_j(t) \quad \text{with } j \in \arg \max_{l \in N} f_l(t),$$

randomizing in case multiple rules yield the maximum average profit. Note that rule revision in this contribution is analogous to the *Imitate-the-Best-Max* rule which is discussed in the context of action revision above. Both NB and SF are satisfied.

Example 2. Under the *Imitate-the-Best-Max with Sampling* criterion (see also [Alós-Ferrer and Weidenholzer 2008](#) for a similar rule in a spatial context) a firm observes a subset of firms in the industry when revising its rule, with all firms having an equal probability of being sampled. Let $m_i(t)$ with $\emptyset \subsetneq m_i(t) \subseteq N$ denote the information sample of firm i after the Cournot competition in period t . A revising firm then adopts the behavioral rule that has yielded the highest average profit within its information sample. Formally, firm i uses

$$\rho_i(t+1) = \rho_j(t) \quad \text{with } j \in \arg \max_{l \in m_i(t)} f_l(t),$$

randomizing if multiple rules yield the maximum average profit in the information sample. This criterion satisfies the NB and SF principles.

A special case of this criterion is obtained when a firm observes only itself and one randomly selected peer, which corresponds to the *Pairwise-Comparison* rule under strong selection ([Traulsen et al., 2007](#)).

Example 3. Under the *Experimental* criterion, as introduced by [Juang \(2002\)](#), firms randomly select from the rules that are in use. Formally, a revising firm i uses $\rho_i(t+1) = \rho_j(t)$ with $j \in N$. Again, this criterion satisfies both NB and SF.

Example 4. Under the *Imitate-if-Better* criterion discussed in [Duersch et al. \(2012\)](#), a revising firm i keeps its current rule $\rho_i(t)$ unchanged if its average profit is maximal in the industry. Otherwise, it switches to a rule that is used by one of the firms with strictly higher average profit. Ties break randomly. Formally,

$$\begin{aligned} \rho_i(t+1) &= \rho_i(t) \quad \text{if } i \in \arg \max_{l \in N} f_l(t); \\ \rho_i(t+1) &= \rho_j(t) \quad \text{with } j \in N \text{ such that } f_j(t) > f_i(t), \text{ otherwise.} \end{aligned}$$

Note that this criterion fulfills NB but violates SF, as firms with the maximum average profit stick to their current rule rather than randomizing among all rules that yield maximal profit.¹⁹

Our framework allows for heterogeneity among firms in rule revision. Thus, not all firms have to follow the same criterion, but rather could apply different criteria belonging to this class. For instance, some may follow the *Imitate-the-Best-Max* criterion, others may adopt the *Experimental* criterion, while still others may employ alternative criteria, as long as they respect the two basic principles.

Further, note that different criteria may have different informational requirements. While the *Imitate-the-Best-Max* criterion requires firms to observe everybody else's average profit, other criteria may require only information on a subset of firms in the industry (e.g. *Imitate-the-Best-Max with Sampling* criterion) or no information about average profits at all (e.g. *Experimental* criterion).

2.3 Mistakes and Long-run equilibria

The dynamics of action- and rule- revision gives rise to a time-homogeneous Markov process $\{\omega(t)\}_{t \geq M}$ over a finite state space $\mathcal{S}$, where

$$\omega(t) = \left(\left(\mathbf{q}(r), \boldsymbol{\rho}(r), \tilde{\boldsymbol{\tau}}(r) \right) \right)_{r=t-M+1}^t.$$

This process is known as the *unperturbed dynamics*, denoted by P . An *absorbing set* of this process is a minimal subset Ω of the state space such that, once the dynamics enters Ω , the probability of leaving it is zero. We denote the set of absorbing sets by Z and use $z = |Z|$ to denote the number of absorbing sets.

Note that once the unperturbed process reaches an absorbing set, it will stay there forever. In order to assess which kind of rules and actions will emerge in the long run, we consider a noisy version of the revision process where firms occasionally make mistakes. Specifically, there are two types of mistakes: *action mistakes* and *rule mistakes*. Action mistakes and rule mistakes are assumed to be independent of one another and independent across firms and over time. Action mistakes happen with fixed probability $\varepsilon \in (0, 1)$.

¹⁹Of course, this could be fixed by adapting the *Imitate-if-Better* criterion by imposing that a firm with maximal profit randomizes among all rules that also yield maximal profit.

When such a mistake occurs an updating firm i ignores the prescription of her rule and selects a quantity from Γ at random, according to a pre-specified distribution with full support. Likewise with probability ε^η a revising agent is subject to a mistake in rule choice, ignores her rule revision prescription and adopts a rule from $\mathcal{R}$ at random. The parameter η captures the relative frequency of rule mistakes compared to action mistakes. We focus on $\eta \geq 1$, so that action mistakes occur more frequently than rule mistakes.²⁰ This assumption captures the intuitive idea that firms are less prone to mistakes when choosing rules for action revision than when applying these rules to choose specific actions.

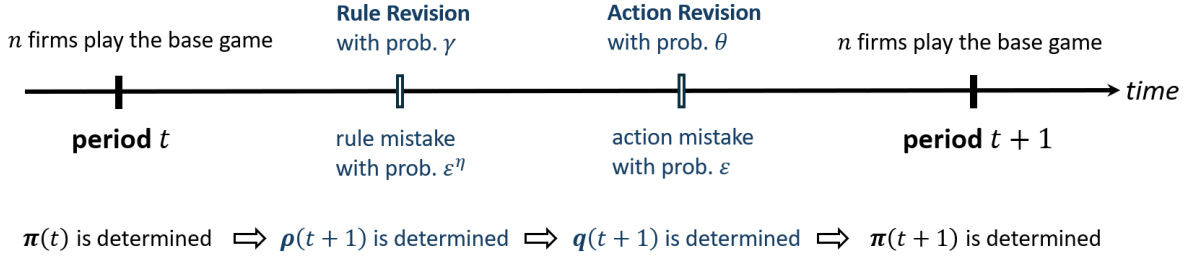


Figure 1: Illustration of the perturbed dynamics with action and rule revision.

The process with mistakes is referred to as the *perturbed dynamics*, denoted by P^ε . For each $\varepsilon > 0$, P^ε is an irreducible and aperiodic Markov chain and has a unique invariant distribution $\mu(\varepsilon)$. It is a well-established result that $\mu^* = \lim_{\varepsilon \rightarrow 0} \mu(\varepsilon)$ exists and is an invariant distribution of the unperturbed dynamics P (Freidlin and Wentzell, 1984; Kandori et al., 1993; Young, 1993). The states in the support of μ^* are called *long-run equilibria* (LRE) and provide a prediction for the long-run behavior of the dynamic process for vanishing ε .

We identify LRE using an algorithm due to Freidlin and Wentzell (1984). We denote the transition cost in terms of the minimum number of mistakes required for the (direct) transition from one absorbing set Ω to another Ω' by $C(\Omega, \Omega') > 0$. An Ω -tree is a directed tree rooted into the absorbing set Ω where the nodes of this tree comprise all absorbing sets. The cost of this tree is given by the sum of the transition cost of each edge. Then an absorbing set Ω is LRE if and only if there exists an Ω -tree with minimal cost among all trees.²¹ The set of LRE is denoted by Z' .

²⁰See also Juang (2002) and Jackson and Watts (2002) for contributions where mistakes in different components of agents' strategies arrive at different rates.

²¹Strictly speaking, LRE is defined at the state level: all states in Ω are LRE if and only if the Ω -tree

3 Results

We are now able to present and discuss our results. To facilitate the exposition, we introduce the following notation. Let $\text{mon}(q, \rho)$ denote the set of states in which all firms employ rule ρ and produce quantity q over M periods, and hence earn the same average profit $\pi(q, nq)$. Formally,

$$\text{mon}(q, \rho) := \left\{ \left((\mathbf{q}^{(r)}, \boldsymbol{\rho}^{(r)}, \tilde{\boldsymbol{\tau}}^{(r)}) \right)_{r=1}^M \in \mathcal{S} \mid \mathbf{q}^{(r)} = (q, \dots, q), \boldsymbol{\rho}^{(r)} = (\rho, \dots, \rho), \text{ for all } r \right\}.$$

Our first insight concerns the rules used by firms in absorbing sets of the unperturbed dynamics.

Lemma 1. *All firms use the same behavioral rule in any absorbing set.*

Intuitively, for any state where firms use different behavioral rules, the *Survival-of-the-Fittest* principle implies that there exists positive probability that all firms adopt the same rule. Given any such state, the *No-Birth* principle ensures that no other rule can emerge. We remark that the reasoning does not depend on the strategic nature of the base game or the set of available behavioral rules. In this sense, Lemma 1 can easily be adopted to a more general analysis of rule evolution.

Our next result characterizes convergence of our process when all firms use BR .

Lemma 2. *From any state where all firms use BR the process converges to $\text{mon}(q^N, BR)$ with probability one.*

We remark that if all firms choose their quantities according to BR , then by *No-Birth* nobody will switch to IM . From there it follows that our process globally converges to the set of states where all firms choose the Cournot Nash quantity q^N . In the proof we establish that the Cournot game under consideration is a finite game with strict strategic substitutes for which the best response dynamics with inertia in action revision converges to the unique Cournot Nash equilibrium. This result is an application of Theorem 2 in [Kukushkin \(2004\)](#).

Given the previous results, we are able to characterize the set of absorbing sets.

has minimal cost among all trees; otherwise, none of them are. With a slight abuse of terminology, we describe such an absorbing set Ω as LRE.

Proposition 1. *All firms use the same behavioral rule and produce the same quantity in any absorbing set. The set of absorbing sets is given by*

$$Z = \left\{ \text{mon}(q, IM) \mid q \in \Gamma \right\} \cup \left\{ \text{mon}(q^N, BR) \right\}.$$

To appreciate the logic that underlies the proof of Proposition 1, note that by Lemma 1, there are two possible types of absorbing sets: (i) sets of states in which all firms adopt IM , and (ii) sets of states in which all firms adopt BR . Note that because of the *No-Birth* principle in neither case can a new rule be introduced into the population. For case (i) note that, because of the *Survival-of-the-Fittest* principle, in the absence of mistakes no new quantity emerges when all firms adopt IM and choose the same quantity. For case (ii), Lemma 2 establishes convergence to the Cournot Nash equilibrium.

We now proceed to present a series of lemmata that characterize the transitions between absorbing sets. For the sake of exposition we focus our discussion on transitions that are at the heart of our results and relegate the proof to Appendix A.2.

The first lemma is based on Vega-Redondo (1997).

Lemma 3. *Consider absorbing sets where all firms use IM . For any $q \neq q^W$ the transition costs are characterized by*

$$C\left(\text{mon}(q, IM), \text{mon}(q^W, IM)\right) = 1$$

and

$$C\left(\text{mon}(q^W, IM), \text{mon}(q, IM)\right) > 1.$$

The underlying idea is that the Walrasian quantity q^W yields a strict relative-payoff advantage over any alternative quantity (Schaffer, 1988; Vega-Redondo, 1997). More formally, for all $q \neq q^W$ and all $1 \leq m < n$,

$$p(mq^W + (n-m)q)q^W - c(q^W) > p(mq^W + (n-m)q)q - c(q). \quad (1)$$

Thus, in an industry of imitators reaching the Walrasian equilibrium requires only a single action mistake, while leaving it requires more mistakes.

The next lemma generalizes the previous insight by arguing that in fact, in addition to the Walrasian quantity, other deviations away from symmetric output profiles may also

improve the relative payoff advantage of a firm. To this end, denote the difference in profits between a single firm producing q' and all other firms producing q by

$$\Delta(q, q') = p(q' + (n-1)q)(q' - q) + c(q) - c(q').$$

Note that if $\Delta(q, q') \geq 0$, then a firm deviating from the output profile of $\text{mon}(q, IM)$ to q' will enjoy a relative payoff advantage and may be imitated. Consequently denote by $D(q) = \{q' \in \mathbb{R}_+ | \Delta(q, q') \geq 0\}$ the set of quantities that will be imitated with positive probability starting from a profile where everybody chooses q .

The following lemma establishes important properties of the set $D(q)$ and discusses the implications for transition costs.

Lemma 4. *i) The set $D(q)$ is characterized as follows*

$$D(q) = \begin{cases} [q, h(q)] & \text{if } q < q^W \\ [\ell(q), q] & \text{if } q > q^W \\ \{q^W\} & \text{if } q = q^W \end{cases}$$

where $\ell(q) < q^W < h(q)$. ii) For any absorbing set $\text{mon}(q, IM)$ with $q \neq q^W$, we have

$$C(\text{mon}(q, IM), \text{mon}(q', IM)) = 1$$

for all $q' \in \Gamma \cap (D(q) \setminus \{q\})$.

In essence, when everybody chooses a quantity below the Walrasian quantity $q < q^W$, then certain quantity-increasing mistakes will be copied and when everybody chooses a quantity above the Walrasian quantity $q > q^W$, then certain quantity-decreasing mistakes will be followed. Figure 2 illustrates this point by plotting the function $\Delta(q, q')$ and the set $D(q)$ for a quantity below the Walrasian quantity and one above the Walrasian quantity.

The next lemma characterizes transitions between absorbing sets where firms uniformly use IM and the absorbing set where firms uniformly use BR .

Lemma 5. *Consider the absorbing sets $\text{mon}(q^N, BR)$, $\text{mon}(q^N, IM)$, and $\text{mon}(q^W, IM)$. Then*

$$C(\text{mon}(q^N, BR), \text{mon}(q^N, IM)) = C(\text{mon}(q^W, IM), \text{mon}(q^N, BR)) = \eta \geq 1.$$

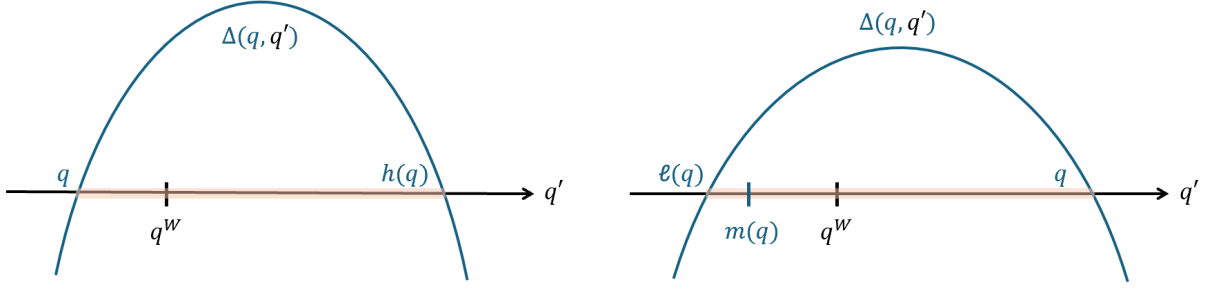


Figure 2: Illustration of the function $\Delta(q, q')$. The horizontal axis plots the variable q' . For any fixed $q < q^W$ ($q > q^W$), the curve in the left (right) panel depicts the value of $\Delta(q, q')$. The orange shade on the axis depicts the set $D(q)$. The definition of $m(q)$ shown in the right panel is provided in Appendix A.

Thus, both of these transitions only require one rule mistake (at the cost of η). To see this, start with the transition from $\text{mon}(q^N, BR)$ to $\text{mon}(q^N, IM)$. Assume that firm i makes a rule mistake and switches to IM . Since there is no action mistake, firm i continues to play q^N , yielding the same profit as all other firms. Because of the *Survival-of-the-Fittest* principle there is a positive probability that all other firms come to adopt IM and the process reaches $\text{mon}(q^N, IM)$. For the transition from $\text{mon}(q^W, IM)$ to $\text{mon}(q^N, BR)$ there are actually two alternative paths which both only require one mistake in rule choice: For the first path consider $\text{mon}(q^W, IM)$ and assume (due to inertia in rule choice) firms have not revised their rules for the past x periods ($1 \leq x < M$). From there consider a rule mistake by firm i which immediately chooses the best response quantity which we denote by $\tilde{q} = b_i((n-1)q^W)$.²² Now observe that the average past performance of i 's rule is $f_i = \pi(\tilde{q}, (n-1)q^W + \tilde{q}) > \pi^W$. In contrast, for all other firms the average past performance of their rule is given by $f_j = \frac{x\pi^W + \pi(q^W, (n-1)q^W + \tilde{q})}{x+1}$. For sufficiently large x , we have $f_i > f_j$ and so other firms will adopt the BR rule with positive probability. Regarding the second path, consider $\text{mon}(q^W, IM)$ and assume firm i makes a mistake and switches its rule to BR but (due to inertia in action revision) does not adjust its quantity yet. Now both IM and BR yield the same past average performance and due to *Survival-of-the-Fittest* with positive probability all other firms switch to BR . Note that while the first path utilizes inertia in rule revision, the second one is based on inertia in quantity revision.²³

²²Without loss of generality, take $\tilde{q} \in \Gamma$. More details are provided in Appendix A.

²³Thus our results would also hold under alternative revision processes that model inertia in a more general way than in the present exposition. Moreover, this indicates that our results do not hinge on intertemporal comparisons of payoffs, which in themselves may introduce a force that pushes away from

With these insights at hand, we are able to state our main result.

Proposition 2. *The set of long-run equilibria is characterized as follows.*

(i) *If $\eta > 1$, then*

$$Z' = \text{mon}(q^N, BR) \cup \text{mon}(q^W, IM),$$

(ii) *If $\eta = 1$, there exist thresholds $\underline{q}$ and $\bar{q}$ satisfying $\underline{q} \leq q^N < q^W \leq \bar{q}$, such that*

$$Z' = \text{mon}(q^N, BR) \cup \left(\bigcup_{q \in \Gamma \cap [\underline{q}, \bar{q}]} \text{mon}(q, IM) \right).$$

When rule mistakes occur less frequently than action mistakes, $\eta > 1$, the logic is as follows. By lemmata 3 and 5 we can exhibit $\text{mon}(q^N, BR)$ and $\text{mon}(q^W, IM)$ -trees of cost $z + \eta - 2$ where z is the number of absorbing sets. Because of the *No-Birth* principle, we can only move between absorbing sets where firms choose different rules with a rule mistake (at cost $\eta > 1$). It follows that these two trees are of minimum cost. Moreover, any other tree has to include a branch going out of $\text{mon}(q^N, BR)$ at cost η and a branch going out of $\text{mon}(q^W, IM)$ at cost strictly larger than 1. It follows that $\text{mon}(q^N, BR) \cup \text{mon}(q^W, IM)$ is the unique LRE.

When rule mistakes occur at the same frequency as action mistakes, $\eta = 1$, it is still true that there exist $\text{mon}(q^N, BR)$ and $\text{mon}(q^W, IM)$ -trees of minimum cost $z - 1$. However, there now exist several other $\text{mon}(q, IM)$ -trees with minimal cost. For instance, consider $\tilde{q} = h(q^N) > q^W$. Because of lemma 4 we can move from $\text{mon}(q^N, IM)$ to $\text{mon}(\tilde{q}, IM)$ with one mistake. It follows there exists a $\text{mon}(\tilde{q}, IM)$ -tree the cost of which is also equal to $z - 1$.²⁴ Using the same and related ideas, we can use the characterization provided in Lemma 4 to identify the entire set of long run equilibria.

Remark. *The exact values of $\underline{q}$ and $\bar{q}$ depend on the number of firms in the industry. Specifically, if $n = 2$ (a duopoly), then $\underline{q} = q^N$ and $\bar{q} = h(q^N)$; if $n \geq 3$, and the grid of feasible output Γ is sufficiently fine, then $\underline{q} = 0$ and $\bar{q} = h(0)$. The detailed proof is relegated to Appendix A.3.*

the Walrasian equilibrium (see Alós-Ferrer 2004).

²⁴In this tree all other absorbing states are connected to $\text{mon}(q^W, IM)$ at the cost of one mistake each and $\text{mon}(q^W, IM)$ is connected to $\text{mon}(q^N, BR)$ at the cost of one mistake.

To facilitate understanding of our results, we present the following example.

Example 5. Consider $n = 4$ firms, an inverse demand function of $p(Q) = \max\{90 - Q, 0\}$, and a cost function of $c(q) = \frac{1}{2}q^2$. Let the grid of feasible output be $\Gamma = \{0, 1, \dots, 90\}$. The Nash quantity is $q^N = 15$, while the Walrasian quantity is $q^W = 18$. The collusive quantity, defined as $q^C := \arg \max_{q \in \mathbb{R}_+} \pi(q, nq)$, is $q^C = 10$. If $\eta > 1$, then $Z' = \text{mon}(18, IM) \cup \text{mon}(15, BR)$ and if $\eta = 1$, then $Z' = \left(\bigcup_{q \in \{0, 1, \dots, 60\}} \text{mon}(q, IM) \right) \cup \text{mon}(15, BR)$.

Several intriguing points stand out. First, $\text{mon}(q^N, BR)$ is always LRE even though imitators earn higher relative profits than best responders in mixed profiles (Schipper, 2009). Since all firms respect the *Survival-of-the-Fittest* principle when revising their behavioral rules, the rule that yields higher profit tends to be attractive. Nevertheless, our results show that states in which firms uniformly adopt the *BR* rule can still emerge in the long run when firms may revise their rules. Second, when rule mistakes occur at the same frequency as action mistakes ($\eta = 1$), our findings reveal a sharp contrast to the unique Walrasian outcome predicted in Vega-Redondo (1997) under uniform adoption of the *IM* rule. Specifically, the introduction of rule revision leads to a broad spectrum of long-run output. Firms' output can persist strictly below q^N when the industry comprises at least three firms. Notably, the output level may fall to the *collusive* quantity, in which case the industry achieves maximal efficiency.

4 Conclusion

This paper highlights the implications of endogenous rule revision in the Cournot oligopoly. Although our analysis has been developed in the context of Cournot oligopolies, the mechanisms underlying our results are not confined to this particular setting. Rather, they rely on the general property of quasi-submodularity, which characterizes a broad class of aggregative games (Corchón, 1994). Prominent examples within this category include Cournot oligopoly, the tragedy of the commons, rent-seeking games, and common-pool resource games (Schipper, 2004; Alos-Ferrer and Ania, 2005). In Cournot terms, quasi-submodularity implies that a firm's payoff advantage of a higher output over a lower one preserves as the aggregate quantity decreases, which can be viewed as a form of strategic substitute. Crucially, extending the results to aggregative games with strict

quasi-submodularity does not alter the economic implications of our findings. Appendix B provides an extensive discussion.

One promising avenue for future research is to examine the robustness of our results by incorporating the costs of behavioral rules (Conlisk, 1980; Droste et al., 2002; Oprea, 2020). Best response, being more computationally demanding, may entail higher implementation costs, whereas imitation learning, which relies on information about both actions and payoffs, may involve greater information costs. Another important direction is to investigate alternative principles of rule revision, particularly those of an uncoupled or reinforcement nature (Fatas et al., 2024). Such principles rely solely on personal information, in contrast to our principles grounded in imitative heuristics. Studying such models may provide valuable insights into the evolution of rules in environments where the rules adopted by others are unobservable or difficult to infer.

A Appendix: Main Proofs

A.1

Proof of Lemma 1. Consider any state in the state space, $\omega \in \mathcal{S}$. After the base game is played, suppose that all firms receive opportunities to revise their behavioral rules. Let i^* be a firm with the highest average profit in the industry, i.e., $i^* \in \arg \max_{l \in N} f_l$. Note that, under the *Survival-of-the-Fittest* principle, each firm with positive probability switches to the rule employed by a firm with average profit no less than its own. In particular, each firm may switch to the rule currently employed by firm i^* . In that case, all firms adopt the same behavioral rule in the next period. Further, by the principle of *No-Birth*, no new rule emerges thereafter. Therefore we conclude that from any initial state, there exists an unperturbed revision path leading to a state in which all firms adopt the same behavioral rule. $\square$

Proof of Lemma 2. We first show that a strictly decreasing inverse demand function $p(\cdot)$ implies that the Cournot oligopoly with a finite strategy set satisfies the condition of *strict strategic substitutes* in the sense of Kukushkin (2004).²⁵ Specifically, for any two feasible output levels $q^{(1)}$ and $q^{(2)}$ with $q^{(2)} > q^{(1)}$ and associated total industry output levels $Q^{(2)} > Q^{(1)}$, a Cournot oligopoly satisfies the strict strategic substitutes condition if

$$\pi(q^{(2)}, Q^{(2)}) \geq \pi(q^{(1)}, Q^{(2)}) \quad \text{implies} \quad \pi(q^{(2)}, Q^{(1)}) > \pi(q^{(1)}, Q^{(1)}).$$

To see this, note that the strictly decreasing $p(\cdot)$ implies $p(Q^{(1)}) > p(Q^{(2)})$. Hence, for any $q^{(2)} > q^{(1)}$, we have

$$p(Q^{(1)})(q^{(2)} - q^{(1)}) > p(Q^{(2)})(q^{(2)} - q^{(1)}). \tag{A.1}$$

Now, suppose that $\pi(q^{(2)}, Q^{(2)}) \geq \pi(q^{(1)}, Q^{(2)})$. By the definition of profit, it follows that

$$p(Q^{(2)})(q^{(2)} - q^{(1)}) \geq c(q^{(2)}) - c(q^{(1)}). \tag{A.2}$$

²⁵This condition is encompassed by the definition of *strict quasi-submodularity*. See Schipper (2009).

Combining A.1 and A.2, we obtain $p(Q^{(1)})(q^{(2)} - q^{(1)}) > c(q^{(2)}) - c(q^{(1)})$, which implies that $\pi(q^{(2)}, Q^{(1)}) > \pi(q^{(1)}, Q^{(1)})$.

Now, consider any state $\omega \in \mathcal{S}$ in which all firms employ BR and assume that (due to inertia) in each period, only one firm adjusts its output. Then, based on Theorem 2 in Kukushkin (2004), after a finite number of periods, the sequence of output profiles converges to equilibrium. Moreover, under our assumptions, the Cournot Nash equilibrium is unique. Hence, all firms eventually produce the Nash quantity q^N , and no firm has an incentive to deviate thereafter. Therefore, we conclude that from any state where all firms use BR the unperturbed dynamics P converges globally to $\text{mon}(q^N, BR)$. $\square$

Proof of Proposition 1. First, it is straightforward that $\text{mon}(q, IM)$ for any $q \in \Gamma$, as well as $\text{mon}(q^N, BR)$, are absorbing sets. It remains to show that the unperturbed dynamics P converges to one of these sets.

According to Lemma 1, starting from any state in the state space, after finite periods, the dynamics will reach a state in which all firms adopt either IM or BR . On one hand, if all firms adopt IM , then with positive probability, they switch to the output that yields the highest profit in the current period. As a result, all firms produce the same quantity from then on and eventually the dynamics reaches some $\text{mon}(q, IM)$, $q \in \Gamma$. On the other hand, if all firms adopt BR , then as established in Lemma 2, all firms eventually produce the Nash quantity and the dynamics reaches $\text{mon}(q^N, BR)$. $\square$

A.2

Proof of Lemma 4. To characterize the set $D(q) := \{q' \in \mathbb{R}_+ \mid \Delta(q, q') \geq 0\}$, we first illustrate the following properties of the function $\Delta(q, q') := p(q' + (n-1)q)(q' - q) + c(q) - c(q')$. Specifically, (i) for $\forall q$, $\Delta(q, q) = 0$; (ii) $\forall q \in \mathbb{R}_+ \setminus \{q^W\}$, $\Delta(q, q^W) > 0$; (iii) $\forall q \in [0, q^W)$, $\Delta(q, q')$ is strictly concave in $q' \in [q, Q_{\max} - (n-1)q]$.

Note that (i) follows directly from the definition of Δ , and (ii) is an immediate implication of Lemma 3. Regarding (iii), note that

$$\frac{\partial^2 \Delta(q, q')}{\partial q'^2} = p''(q' + (n-1)q)(q' - q) + 2p'(q' + (n-1)q) - c''(q'). \quad (\text{A.3})$$

Because $p'' \leq 0$, $p' < 0$, and $c'' \geq 0$, A.3 implies (iii).

We now show that the set $D(q)$ is characterized as

$$D(q) = \begin{cases} [q, h(q)] & \text{if } q < q^W \\ [\ell(q), q] & \text{if } q > q^W \\ \{q^W\} & \text{if } q = q^W \end{cases}$$

where $\ell(q) < q^W < h(q)$. We proceed by analyzing the three cases in turn.

Case 1 For any fixed $q \in [0, q^W)$ properties (i)-(iii) imply that the equation $\Delta(q, q') = 0$ has exactly two solutions. One is the trivial solution $q' = q$; the other, denoted by $q' = h(q)$, satisfies $h(q) > q^W$ and $h(q) < Q_{max} - (n-1)q$. Hence, the function $h : [0, q^W) \rightarrow (q^W, Q_{max})$ is well defined. As illustrated by the left panel of Figure 2, $\Delta(q, q') \geq 0$ if and only if $q' \in [q, h(q)]$. Therefore, we conclude that $D(q) = [q, h(q)]$ for $q < q^W$, where $h(q) > q^W$.

Case 2 For any fixed $q > q^W$, properties (i)-(iii) imply that $\Delta(q, q') < 0$ for all $q' > q$. Hence, $D(q) \subseteq [0, q]$. To further characterize $D(q)$ we denote by $q' = m(q) \geq 0$ the unique solution of $p(q' + (n-1)q) - c'(q) = 0$.²⁶ Since $q > q^W$, it is obvious that $m(q) < q^W$. See the right panel of Figure 2 for an illustration.

Note that for all $q' \in [m(q), q]$,

$$\Delta(q, q') \geq c'(q')(q' - q) + c(q) - c(q') = (c'(\xi) - c'(q'))(q - q') \geq 0, \quad (\text{A.4})$$

where $q' \leq \xi \leq q$. Thus, $[m(q), q] \subseteq D(q)$, for all $q > q^W$. Further, note that when $q' \in [0, m(q)]$,

$$\frac{\partial \Delta(q, q')}{\partial q'} = p'(q' + (n-1)q)(q' - q) + p(q' + (n-1)q) - c'(q') > 0. \quad (\text{A.5})$$

Thus, $\Delta(q, q')$ strictly increases in $[0, m(q)]$, and hence there exists unique $\ell(q) \in [0, m(q)]$ such that $\Delta(q, q') \geq 0$ if and only if $q' \in [\ell(q), q]$. Specifically, if $\Delta(q, 0) \leq 0$, $\ell(q) \in$

²⁶Without loss of generality, we assume that $p((n-1)q) - c'(0) > 0$. If $p((n-1)q) - c'(0) \leq 0$, it is clear that $D(q) = [0, q]$.

$[0, m(q)]$ satisfies $\Delta(q, \ell(q)) = 0$; otherwise, if $\Delta(q, 0) > 0$, then $\ell(q) = 0$. It follows that for any $q > q^W$, $D(q) = [\ell(q), q]$, where $\ell(q) < q^W$.²⁷

Case 3 Lemma 3 implies $D(q^W) = \{q^W\}$.

Finally, Lemma 4 ii) follows directly from Lemma 4 i). □

Proof of Lemma 5. We first show that a single rule mistake can induce a transition from $\text{mon}(q^N, BR)$ to $\text{mon}(q^N, IM)$. The reasoning proceeds as follows. Consider any state in $\text{mon}(q^N, BR)$. From there suppose that firm i revises its rule and mistakenly adopts IM . Accordingly, firm i updates its output by imitation and hence still produces the Nash quantity q^N . Since all firms continue to choose the Nash quantity, they share the same average profit π^N , regardless of the length of the evaluation window for each firm. As a result, BR and IM yield the same average performance. By the *Survival-of-the-Fittest* principle, all other firms may subsequently adopt IM as well. Regarding the output level, either by imitation or inertia, all firms continue to produce q^N . Therefore, we conclude that $C(\text{mon}(q^N, BR), \text{mon}(q^N, IM)) = \eta$.

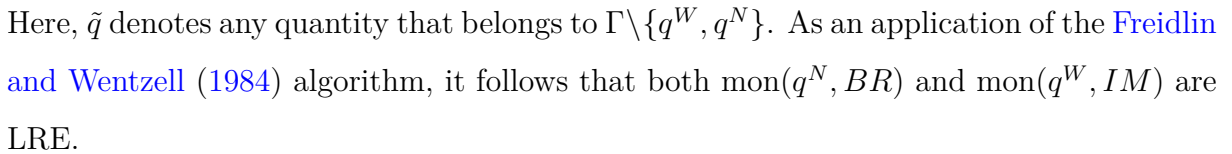
Next, we show that one rule mistake can lead to the transition from $\text{mon}(q^W, IM)$ to $\text{mon}(q^N, BR)$. There are actually two possible revision paths. For the first path, starting from any state in $\text{mon}(q^W, IM)$, consider that the dynamics reaches a state in which no firm has revised its rule for the past x periods ($1 \leq x < M$) due to inertia in rule choice. From there consider a rule mistake by firm i which immediately chooses a best response quantity, denoted by $\tilde{q} \in \Gamma$. Then two cases arise. On one hand, if q^W is not a best response to the Walrasian equilibrium, the average past performance of i 's current rule (i.e., BR) exceeds i 's previous profit under the Walrasian outcome. That is, $f_i = \pi(\tilde{q}, (n-1)q^W + \tilde{q}) > \pi^W$. Meanwhile, for all other firms, the average past performance of their rule (i.e., IM) is given by $f_j = \frac{x\pi^W + \pi(q^W, (n-1)q^W + \tilde{q})}{x+1}$. For sufficiently large x , we have $f_i > f_j$ and so other firms will adopt the BR rule with positive probability. On the other hand, if q^W is a best response to the Walrasian equilibrium (which may occur when the grid Γ is sufficiently coarse), then firm i may still choose the Walrasian quantity. Since all firms continue to choose the Walrasian quantity, they share the same average profit π^W for any x . As a result, IM and BR yield the same average performance and

²⁷Unlike in case 1, in case 2, $\Delta(q, q')$ is not necessarily strictly concave in $q' \in [0, q]$, and therefore the proof cannot proceed in the same way as in case 1.

Regarding the second path, consider $\text{mon}(q^W, IM)$ and assumes that a firm mistakenly switches to BR but (due to inertia in action choice) does not adjust its quantity yet. Note that IM and BR yield the same average performance and because of the *Survival-of-the-Fittest* principle with positive probability all other firms switch to BR . In light of the two paths mentioned above, by Lemma 2, the dynamics finally converges to $\text{mon}(q^N, BR)$. We further conclude that $C(\text{mon}(q^W, IM), \text{mon}(q^N, BR)) = \eta$.

Proof of Proposition 2. With Lemmata 4 and 5 established, we now characterize the long-run equilibria.

Next, we show that there exists a $\text{mon}(q^W, IM)$ -tree and a $\text{mon}(q^N, BR)$ -tree, each with cost equal to $z - 2 + \eta$. Specifically, the corresponding trees are illustrated below:



26

node of any $(\tilde{q}, IM)$ -tree, the tree must include an edge starting from $\text{mon}(q^W, IM)$ and another starting from $\text{mon}(q^N, BR)$. The cost of the former is strictly greater than 1, and the cost of the latter is no less than η . Therefore, the cost of any $\text{mon}(\tilde{q}, IM)$ -tree with $\tilde{q} \neq q^W$ is strictly greater than $z - 2 + \eta$. As an application of the algorithm in [Freidlin and Wentzell \(1984\)](#), it follows that $\text{mon}(\tilde{q}, IM)$ is not LRE, for all $\tilde{q} \neq q^W$. We further conclude that when $\eta > 1$, $Z' = \text{mon}(q^N, BR) \cup \text{mon}(q^W, IM)$.

Case 2 Consider the case where $\eta = 1$. Then, for any absorbing set Ω and any Ω -tree, the cost of the tree must be at least $z - 1$. We further demonstrate that there exists a $\text{mon}(q^W, IM)$ -tree, a $\text{mon}(q^N, BR)$ -tree and a $\text{mon}(q^N, IM)$ -tree, each with cost equal to $z - 1$. Specifically, the corresponding trees are shown below:

$$\begin{array}{ccccc} \text{mon}(q^W, IM) & \xleftarrow{1} & \text{mon}(q^N, IM) & \xleftarrow{1} & \text{mon}(q^N, BR) \\ \uparrow 1 & & & & \\ \text{mon}(\tilde{q}, IM) & & & & \end{array}$$

$$\begin{array}{ccccc} \text{mon}(q^N, BR) & \xleftarrow{1} & \text{mon}(q^W, IM) & \xleftarrow{1} & \text{mon}(q^N, IM) \\ & & \uparrow 1 & & \\ & & \text{mon}(\tilde{q}, IM) & & \end{array}$$

$$\begin{array}{ccccc} \text{mon}(q^N, IM) & \xleftarrow{1} & \text{mon}(q^N, BR) & \xleftarrow{1} & \text{mon}(q^W, IM) \\ & & & & \uparrow 1 \\ & & & & \text{mon}(\tilde{q}, IM) \end{array}$$

Here, $\tilde{q}$ denotes any quantity that belongs to $\Gamma \setminus \{q^W, q^N\}$. From the [Freidlin and Wentzell \(1984\)](#) algorithm, it follows that $\text{mon}(q^W, IM)$, $\text{mon}(q^N, BR)$, and $\text{mon}(q^N, IM)$ are LRE.

Further, by Lemma 4, the transition cost is $C(\text{mon}(q^N, IM), \text{mon}(q', IM)) = 1$ for any $q' \in \Gamma \cap (q^N, h(q^N)]$. Hence, there exists a $\text{mon}(q', IM)$ -tree with cost $x - 1$. Specifically,

$$\begin{array}{ccccc}
\text{mon}(q^N, IM) & \xleftarrow{1} & \text{mon}(q^N, BR) & \xleftarrow{1} & \text{mon}(q^W, IM) \\
\downarrow 1 & & & & \uparrow 1 \\
\text{mon}(q', IM) & & & & \text{mon}(\tilde{q}, IM)
\end{array}$$

where $\tilde{q}$ denotes any quantity in $\Gamma \setminus \{q', q^W, q^N\}$. The [Freidlin and Wentzell \(1984\)](#) algorithm, implies that $\text{mon}(q, IM)$ is LRE for all $q \in \Gamma \cap [q^N, h(q^N)]$, where $h(q^N) > q^W$. We now define

$$\underline{q} := \min \{q \in \Gamma \mid \text{mon}(q, IM) \text{ is LRE}\}, \quad \bar{q} := \max \{q \in \Gamma \mid \text{mon}(q, IM) \text{ is LRE}\}.$$

Based on the previous analysis, we obtain that $\underline{q} \leq q^N$ and $\bar{q} \geq q^W$. Further, by Lemma 4, $\Delta(\underline{q}, q') \geq 0$ for all $q' \in (\underline{q}, q^N)$, if such q' exists. Hence, a single action mistake can lead to the transition from $\text{mon}(\underline{q}, IM)$ to $\text{mon}(q', IM)$, for all $q' \in \Gamma \cap (\underline{q}, q^N)$. Besides, a single action mistake can also lead to the transition from $\text{mon}(q', IM)$ to $\text{mon}(q^W, IM)$. As an application of Theorem 3 in [Ellison \(2000\)](#), which is referred to as the *tree surgery* argument, we conclude that $\text{mon}(q', IM)$ is LRE for all $q' \in \Gamma \cap (\underline{q}, q^N)$.

Further, Lemma 4 implies that $\Delta(\bar{q}, q') \geq 0$ for all $q' \in (q^W, \bar{q})$, if such q' exists. Following the same reasoning as above, we conclude that $\text{mon}(q', IM)$ is LRE for all $q' \in \Gamma \cap (q^W, \bar{q})$. The proof is thus complete. $\square$

A.3

We first point out additional properties of $\Delta(\cdot, \cdot)$ that are important for understanding our results, continuing the numbering in A.2. Specifically, (iv) when $n = 2$, $\Delta(q, q') = \Delta(q', q)$, for $\forall q, q'$; (v) when $n > 2$, $\Delta(q, q') + \Delta(q', q) > 0$, for $\forall q \neq q'$; (vi) the functions $h(\cdot)$ and $\ell(\cdot)$ defined in Lemma 4 are continuous and strictly decreasing. These properties follow directly from the definitions and the Implicit Function Theorem.

Case 1 Consider the case of $n = 2$ (duopoly). We first show that a single action mistake cannot induce a transition from $\text{mon}(q, IM)$ to $\text{mon}(q', IM)$, for any $q \in \Gamma \cap [q^N, h(q^N)]$ and $q' \notin \Gamma \cap [q^N, h(q^N)]$. To see this, note that by (iv), we have

$$\Delta(q^N, h(q^N)) = \Delta(h(q^N), q^N) = 0. \tag{A.6}$$

Moreover, by (vi), $h(q) < h(q^N)$ for any $q \in (q^N, q^W)$, implying that $D(q) \subset [q^N, h(q^N)]$. Similarly, (vi) also implies that $\ell(q) > \ell(h(q^N))$ for any $q \in (q^W, h(q^N))$. Since Formula (A.6) implies $\ell(h(q^N)) = q^N$, it follows that $\ell(q) > q^N$, so again $D(q) \subset [q^N, h(q^N)]$.

Next, we demonstrate that a single rule mistake cannot lead to a transition from $\text{mon}(q, \text{IM})$ to $\text{mon}(q', \text{IM})$, for any $q \in \Gamma \cap [q^N, h(q^N)]$ and $q' \notin \Gamma \cap [q^N, h(q^N)]$. The reasoning proceeds as follows. Starting from any state in $\text{mon}(q, \text{IM})$, suppose a single rule mistake occurs. Thereafter, along any unperturbed revision path, firms' output must be either q or some iterated best response to q , denoted by $\tilde{q} \in \Gamma$. Since the best response correspondence is a contraction in the duopoly setting (Vives, 1999, Chapter 4), it follows that $\tilde{q} \leq q$.

Moreover, without loss of generality, assume that at least one firm follows *IM* in each period for any unperturbed revision path. Consider any period in which firm 1 produces q_1 and firm 2 produces q_2 . If $q \geq q_1 \geq q_2 \geq q^N$, then one firm will produce q_1 by imitation, or produce either q_1 or q_2 due to inertia. If instead $q \geq q_1 \geq q^N > q_2$, then if firm 1 follows *IM*, it will again produce q_1 via imitation or inertia; if firm 1 follows *BR*, then it will produce q_1 due to inertia or choose a best response to q_2 , which must be no less than q^N . Applying this reasoning recursively, we conclude that if both firms produce q in the initial state, then in each subsequent period, at least one firm's output is no less than q^N . Therefore, the dynamics finally converges to an absorbing set $\text{mon}(q', \text{IM})$, with $q^N \leq q' \leq q \leq h(q^N)$.

Applying the Freidlin and Wentzell (1984) algorithm, we conclude that for any $q' \notin \Gamma \cap [q^N, h(q^N)]$, $\text{mon}(q', \text{IM})$ is not LRE.

Case 2 Consider the case of $n > 2$. Before proceeding to a formal proof, we first analyze Example 5 to build intuition. Note that $h(q^N) = h(15) = 25$. Applying the same reasoning as in A.2, we conclude that $\text{mon}(q', \text{IM})$ is LRE for all $q' = 15, 16, \dots, 25$. Furthermore, since $\ell(25) = \frac{5}{3}$, it follows that $\text{mon}(q', \text{IM})$ is LRE for all $q' = 2, 3, \dots, 25$. Applying this logic recursively, and using the facts that $h(2) = \frac{166}{3}$, $\ell(55) = 0$ (since $\Delta(55, 0) > 0$), and $h(0) = 60$, we ultimately conclude that $\text{mon}(q, \text{IM})$ is LRE for all $q = 0, 1, \dots, 60$.

To establish the proof, we define sequences $\{a_n\} \subset \mathbb{R}_+$ and $\{b_n\} \subset \mathbb{R}_+$ as follows:

$$a_1 = q^N, \quad a_{n+1} = \ell(h(a_n)), \quad \text{and} \quad b_n = h(a_n) \quad \text{for all } n \geq 1.$$

Because of (v), we have

$$\Delta(h(a_n), a_n) = \Delta(a_n, h(a_n)) + \Delta(h(a_n), a_n) > 0. \quad (\text{A.7})$$

Formula (A.7) implies that, if $a_n > 0$, then $\ell(h(a_n)) < a_n$; otherwise, if $a_n = 0$, then $\ell(h(a_n)) = a_n$. That is, $\{a_n\}$ strictly decreases unless it reaches 0. Furthermore, due to (vi), it follows that $\{b_n\}$ strictly increases unless it reaches $h(0)$.

For the decreasing sequence $\{a_n\}$, there are two possible scenarios. First, there exists $k \in \mathbb{N}$ such that $a_k = 0$. Second, for any $n \in \mathbb{N}$, $a_n > 0$. Note that the second scenario can be ruled out. Since $\{a_n\}$ is decreasing and bounded below by zero, it converges to some $a \geq 0$, i.e., $\lim_{n \rightarrow \infty} a_n = a$. Furthermore, given that $h(\cdot)$ and $\ell(\cdot)$ are continuous (as established in (vi)), we obtain

$$a = \lim_{n \rightarrow \infty} a_{n+1} = \ell(h(\lim_{n \rightarrow \infty} a_n)) = \ell(h(a)).$$

Note that $\ell(h(a)) = a$ only when $a = 0$. Hence, the increasing sequence $\{b_n\}$ converges to $h(0)$. Moreover, for any $\epsilon \in (0, \frac{h(0)}{2(n-1)})$,

$$\begin{aligned} \Delta(h(0) - \epsilon, 0) &= -p((n-1)(h(0) - \epsilon))(h(0) - \epsilon) + c(h(0) - \epsilon) - c(0) \\ &= -p((n-1)(h(0) - \epsilon))(h(0) - \epsilon) + c(h(0) - \epsilon) - (c(h(0)) - p(h(0))h(0)) \\ &= (p(h(0)) - p((n-1)(h(0) - \epsilon)))h(0) + \epsilon(p((n-1)(h(0) - \epsilon)) - c'(\xi^\epsilon)) \\ &\geq (p(h(0)) - p((n-1.5)h(0)))h(0) + \epsilon(p((n-1)h(0)) - c'(h(0))) \end{aligned}$$

where $\xi^\epsilon \in [h(0) - \epsilon, h(0)]$. Note that $n \geq 3$ and $h(0) < Q_{max}$, hence it yields

$$p(h(0)) - p((n-1.5)h(0)) > 0.$$

We conclude that $\Delta(h(0) - \epsilon, 0) > 0$ as long as ϵ is sufficiently close to zero, which contradicts with the assumption of the second scenario that $a_n > 0$ for any n .

Provided that the finite strategy set Γ is sufficiently fine, $\{a_n\}_{n=1}^k$ and $\{b_n\}_{n=1}^k$ are contained in Γ , where $a_k = 0$ and $b_k = h(0)$. Applying the transition triggered by one action mistake recursively, we conclude that $\text{mon}(q, IM)$ is LRE for any $q \in \Gamma \cap [0, h(0)]$, when $n > 2$. Moreover, it is evident that no further LRE exists. $\square$

B Appendix: Aggregative Games

We now introduce symmetric aggregative games (Corchón, 1994), where agents choose actions (rather than firms choosing quantities) from a finite strategy set $S \subseteq \mathbb{R}$, forming a lattice.

Definition 3 (*Aggregative games*). A symmetric aggregative game with aggregate g is defined as the tuple $G := (N, S, \tilde{\pi}, g)$, where $g : S^N \rightarrow T$ is a real-valued function that is symmetric and monotone increasing (referred to as the aggregator), and $\tilde{\pi} : S \times T \rightarrow \mathbb{R}$ is a real-valued function such that the individual payoff functions are given by $\pi_i(\mathbf{s}) := \tilde{\pi}(s_i, g(\mathbf{s}))$, for all action profile $\mathbf{s} = (s_1, \dots, s_n) \in S^n$, $i = 1, \dots, n$.

We focus on aggregative games with strict quasi-submodularity. Prominent examples include Cournot oligopolies, the tragedy of the commons, rent-seeking games, and common-pool resource games. Refer to Schipper (2004) and Alos-Ferrer and Ania (2005) for more examples.

Definition 4 (*Quasi-submodularity*). An aggregative game $G = (N, S, \tilde{\pi}, g)$ is strictly quasi-submodular if $\tilde{\pi}$ satisfies the following conditions: for any $s^{(1)} < s^{(2)} \in S$ and $t^{(1)} < t^{(2)} \in T$,

$$\begin{aligned} \tilde{\pi}(s^{(2)}, t^{(1)}) &\leq \tilde{\pi}(s^{(1)}, t^{(1)}) \quad \text{implies} \quad \tilde{\pi}(s^{(2)}, t^{(2)}) < \tilde{\pi}(s^{(1)}, t^{(2)}), \\ \tilde{\pi}(s^{(2)}, t^{(2)}) &\geq \tilde{\pi}(s^{(1)}, t^{(2)}) \quad \text{implies} \quad \tilde{\pi}(s^{(2)}, t^{(1)}) > \tilde{\pi}(s^{(1)}, t^{(1)}). \end{aligned}$$

Analogous to the concept of a Walrasian quantity in the Cournot oligopoly, we now define an *aggregate-taking strategy* for symmetric aggregative games (Possajennikov, 2003).

Definition 5 (*Aggregate-taking strategy*). Assume $G = (N, S, \tilde{\pi}, g)$ is a symmetric aggregative game. A strategy $s^* \in S$ is an aggregate-taking strategy (ATS) if

$$s^* \in \arg \max_{s \in S} \tilde{\pi}(s, g(\vec{s}^*)),$$

where $\vec{s}^*$ denotes the monomorphic action profile in which all agents choose s^* .

With this definition at hand, Formula (1) in the main text can be generalized as follows.

Lemma 6. *We assume that the symmetric aggregative game $G = (N, S, \tilde{\pi}, g)$ is strictly quasi-submodular. Let s^* be an *ATS* of G . For all $s' \neq s^*$ and $1 \leq m < n$, we have*

$$\tilde{\pi}\left(s^*, g\left(\underbrace{(s', \dots, s')}_m, \underbrace{(s^*, \dots, s^*)}_{n-m}\right)\right) > \tilde{\pi}\left(s', g\left(\underbrace{(s', \dots, s')}_m, \underbrace{(s^*, \dots, s^*)}_{n-m}\right)\right).$$

In this section, we assume that game G admits an *ATS*.²⁸ Lemma 6 further shows that the *ATS* is unique whenever G is a strictly quasi-submodular aggregative game. As established by Schipper (2004), it follows that when all agents follow imitation learning, the *ATS* equilibrium in which all agents play the *ATS* is the unique LRE, thereby generalizing the result of Vega-Redondo (1997). We denote by $\text{mon}(\text{ATS}, IM)$ the set of monomorphic states in which all agents adopt *IM* and play the *ATS*.

Moreover, we assume that G admits a unique pure strategy Nash equilibrium, which must be symmetric given the symmetry of G . Following the same reasoning as in Appendix A.1, and relying on quasi-submodularity, we conclude that if all agents adopt best response learning, the dynamics converges globally to this Nash equilibrium. We denote the set of monomorphic states in which all agents use *BR* and play the Nash strategy as $\text{mon}(\text{Nash}, BR)$.

Building on the above analysis, the absorbing sets can be readily identified. We further impose the mild assumption that the Nash strategy *Nash* differs from the *ATS*, and present the results as follows.

Proposition 3. *The set of long-run equilibria is characterized as follows.*

(i) *If $\eta > 1$, then*

$$Z' = \text{mon}(\text{Nash}, BR) \cup \text{mon}(\text{ATS}, IM).$$

(ii) *If $\eta = 1$, then*

$$Z' \supseteq \text{mon}(\text{Nash}, BR) \cup \text{mon}(\text{Nash}, IM) \cup \text{mon}(\text{ATS}, IM).$$

The logic underlying Proposition 3 parallels that in Appendix A.2, and the proof is therefore omitted. These results highlight that, despite the relative advantage of imitators

²⁸Kakutani's fixed point theorem provides sufficient conditions for the existence of an *ATS*.

in strictly quasi-submodular aggregative games ([Schipper, 2009](#)), best-response learning can nonetheless persist in the long run. Moreover, when rule mistakes occur at the same rate as action mistakes, the *ATS* equilibrium is no longer the unique prediction under imitation. We point out that, beyond the Nash and *ATS* equilibria, a range of action profiles may also emerge under imitation learning in the long run, following reasoning similar to that in [Lemma 4](#) and [Appendix A.3](#).

References

- Alós-Ferrer, C., 2004. Cournot versus walras in dynamic oligopolies with memory. *International Journal of Industrial Organization* 22, 193–217.
- Alos-Ferrer, C., Ania, A.B., 2005. The evolutionary stability of perfectly competitive behavior. *Economic Theory* 26, 497–516.
- Alós-Ferrer, C., Ritschel, A., 2021. Multiple behavioral rules in cournot oligopolies. *Journal of Economic Behavior & Organization* 183, 250–267.
- Alós-Ferrer, C., Weidenholzer, S., 2006. Imitation, local interactions, and efficiency. *Economics Letters* 93, 163–168.
- Alós-Ferrer, C., Weidenholzer, S., 2008. Contagion and efficiency. *Journal of Economic Theory* 143, 251–274.
- Alós-Ferrer, C., Weidenholzer, S., 2014. Imitation and the role of information in overcoming coordination failures. *Games and Economic Behavior* 87, 397–411.
- Apesteguia, J., Huck, S., Oechssler, J., 2007. Imitation—theory and experimental evidence. *Journal of Economic Theory* 136, 217–235.
- Apesteguia, J., Huck, S., Oechssler, J., Weidenholzer, S., 2010. Imitation and the evolution of walrasian behavior: Theoretically fragile but behaviorally robust. *Journal of Economic Theory* 145, 1603–1617.
- Bigoni, M., Fort, M., 2013. Information and learning in oligopoly: An experiment. *Games and Economic Behavior* 81, 192–214.

- Blume, L.E., 1993. The statistical mechanics of strategic interaction. *Games and Economic Behavior* 5, 387–424.
- Bosch-Domènech, A., Vriend, N.J., 2003. Imitation of successful behaviour in cournot markets. *The Economic Journal* 113, 495–524.
- Bossan, B., Jann, O., Hammerstein, P., 2015. The evolution of social learning and its economic consequences. *Journal of Economic Behavior & Organization* 112, 266–288.
- Conlisk, J., 1980. Costly optimizers versus cheap imitators. *Journal of Economic Behavior & Organization* 1, 275–293.
- Corchón, L.C., 1994. Comparative statics for aggregative games the strong concavity case. *Mathematical Social Sciences* 28, 151–165.
- Cournot, A.A., 1836. *Researches into the mathematical principles of the theory of wealth*, English edition of *Recherches sur les principes mathématiques de la théorie des richesses*. Kelley, New York.
- Denis, D.J., Denis, D.K., 1995. Performance changes following top management dismissals. *The Journal of finance* 50, 1029–1057.
- Droste, E., Hommes, C., Tuinstra, J., 2002. Endogenous fluctuations under evolutionary pressure in cournot competition. *Games and Economic Behavior* 40, 232–269.
- Dubey, P., Haimanko, O., Zapechelnyuk, A., 2006. Strategic complements and substitutes, and potential games. *Games and Economic Behavior* 54, 77–94.
- Duersch, P., Kolb, A., Oechssler, J., Schipper, B.C., 2010. Rage against the machines: How subjects play against learning algorithms. *Economic Theory* 43, 407–430.
- Duersch, P., Oechssler, J., Schipper, B.C., 2012. Unbeatable imitation. *Games and Economic Behavior* 76, 88–96.
- Ellison, G., 1993. Learning, local interaction, and coordination. *Econometrica* , 1047–1071.
- Ellison, G., 2000. Basins of attraction, long-run stochastic stability, and the speed of step-by-step evolution. *The Review of Economic Studies* 67, 17–45.

- Fatas, E., Morales, A.J., Jaramillo-Gutiérrez, A., 2024. Social aspiration reinforcement learning in cournot games. *Economic Theory* , 1–40.
- Fisher, F.M., 1961. The stability of the cournot oligopoly solution: The effects of speeds of adjustment and increasing marginal costs. *The Review of Economic Studies* 28, 125–135.
- Freidlin, M.I., Wentzell, A.D., 1984. *Random Perturbations of Dynamical Systems*. volume 260. 1st ed., Springer.
- Friedman, D., Huck, S., Oprea, R., Weidenholzer, S., 2015. From imitation to collusion: Long-run learning in a low-information environment. *Journal of Economic Theory* 155, 185–205.
- Fudenberg, D., Levine, D.K., 1998. *The theory of learning in games*. volume 2. MIT Press.
- Hehenkamp, B., Kaarbøe, O.M., 2008. Imitators and optimizers in a changing environment. *Journal of Economic Dynamics and Control* 32, 1357–1380.
- Hofbauer, J., Sigmund, K., 1998. *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge, UK.
- Huck, S., Leutgeb, J., Oprea, R., 2017. Payoff information hampers the evolution of cooperation. *Nature Communications* 8, 15147.
- Huck, S., Normann, H.T., Oechssler, J., 1999. Learning in cournot oligopoly—an experiment. *The Economic Journal* 109, 80–95.
- Huck, S., Normann, H.T., Oechssler, J., 2002. Stability of the cournot process—experimental evidence. *International Journal of Game Theory* 31, 123–136.
- Jackson, M.O., Watts, A., 2002. On the formation of interaction networks in social coordination games. *Games and Economic Behavior* 41, 265–291.
- Juang, W., Sabourian, H., 2021. Rules and Mutation-A Theory of How Efficiency and Rawlsian Egalitarianism/Symmetry May Emerge. Technical Report. Faculty of Economics, University of Cambridge.

- Juang, W.T., 2002. Rule evolution and equilibrium selection. *Games and Economic Behavior* 39, 71–90.
- Kalai, E., Lehrer, E., 1993. Rational learning leads to nash equilibrium. *Econometrica* , 1019–1045.
- Kandori, M., Mailath, G.J., Rob, R., 1993. Learning, mutation, and long run equilibria in games. *Econometrica* , 29–56.
- Kovářík, J., Mengel, F., Romero, J.G., 2018. Learning in network games. *Quantitative Economics* 9, 85–139.
- Kukushkin, N.S., 2004. Best response dynamics in finite games with additive aggregation. *Games and Economic Behavior* 48, 94–110.
- Nax, H.H., Burton-Chellew, M.N., West, S.A., Young, H.P., 2016. Learning in a black box. *Journal of Economic Behavior & Organization* 127, 1–15.
- Newton, J., 2018. Evolutionary game theory: A renaissance. *Games* 9, 31.
- Oechssler, J., Roomets, A., Roth, S., 2016. From imitation to collusion: a replication. *Journal of the Economic Science Association* 2, 13–21.
- Oprea, R., 2020. What makes a rule complex? *American Economic Review* 110, 3913–3951.
- Oyama, D., Sandholm, W.H., Tercieux, O., 2015. Sampling best response dynamics and deterministic equilibrium selection. *Theoretical Economics* 10, 243–281.
- Possajennikov, A., 2003. Evolutionary foundations of aggregate-taking behavior. *Economic Theory* 21, 921–928.
- Schaffer, M.E., 1988. Evolutionarily stable strategies for a finite population and a variable contest size. *Journal of Theoretical Biology* 132, 469–478.
- Schipper, B.C., 2004. Submodularity and the evolution of walrasian behavior. *International Journal of Game Theory* 32, 471–477.
- Schipper, B.C., 2009. Imitators and optimizers in cournot oligopoly. *Journal of Economic Dynamics and Control* 33, 1981–1990.

- Theocharis, R.D., 1960. On the stability of the cournot solution on the oligopoly problem. *The Review of Economic Studies* 27, 133–134.
- Thijssen, J., 2009. Approximate stochastic stability with an application to the survival of walrasian conjectures in cournot oligopoly. Available at SSRN 1425008 .
- Tirole, J., 1988. *The theory of industrial organization*. MIT Press.
- Traulsen, A., Pacheco, J.M., Nowak, M.A., 2007. Pairwise comparison and selection temperature in evolutionary game dynamics. *Journal of Theoretical Biology* 246, 522–529.
- Vega-Redondo, F., 1997. The evolution of walrasian behavior. *Econometrica* , 375–384.
- Vives, X., 1999. *Oligopoly pricing: old ideas and new tools*. MIT Press.
- Wang, X., Zhou, L., McAvoy, A., Li, A., 2023. Imitation dynamics on networks with incomplete information. *Nature Communications* 14, 7453.
- Weisbach, M.S., 1995. Ceo turnover and the firm’s investment decisions. *Journal of Financial Economics* 37, 159–188.
- Young, H.P., 1993. The evolution of conventions. *Econometrica* , 57–84.