Queen's Economics Department Working Paper No. 1042

Cooperation through Imitation

James Bergin
Queen's University

Dan Bernhardt
University of Illinois

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

1-2006

Cooperation through Imitation *

James Bergin
Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, K7L 3N6

Dan Bernhardt
Department of Economics
University of Illinois
1206 S. Sixth Street
Champaign, IL 61820

July 2005

Abstract

This paper characterizes long-run outcomes for broad classes of symmetric games, when players select actions on the basis of average historical performance. Received wisdom is that when agent's interests are partially opposed, behavior is excessively competitive: "keeping up with the Jones' " lowers everyones' welfare. Here, we study the long-run consequences of imitative behavior when agents have sufficiently long memories — and the outcome is dramatically different. Imitation robustly leads to cooperative outcomes (with highest symmetric payoffs) in the long run. This provides a rationale, for example, for collusive cartel-like behavior without collusive intent on the part of the agents.

*Dan Bernhardt gratefully acknowledges support from the National Science Foundation, grant number SES-0317700.

1 Introduction

How does cooperation arise in competitive environments where agents' interests conflict? One answer to this question is that repeated interaction sustains cooperation — cooperation on a period-by-period basis is justified by the need to maintain working relationships. Thus, rational agents with precise knowledge of their environment optimize over time with a web of threats and rewards sustaining cooperative behavior. But, in complex strategic environments, agents may have little faith in their ability to predict the behavior of others and may have limited ability to determine what constitutes a "good decision". In such circumstances, the levels of rationality and knowledge mandated by traditional game theoretic models seem implausible — presupposing, as it does, that individuals have full knowledge of the environment, have the ability to anticipate how others will behave, and are capable of making difficult computations. Concerns such as these have given rise to learning-based models that attempt to provide insight into individual behavior under plausible requirements on agents' ability to reason in and fully understand, the environment in which they function.

This paper takes such an approach and shows how cooperation arises naturally from imitation of successful behavior by agents who have limited information and limited computational ability. It is accepted wisdom that in competitive settings, imitative behavior typically leads to destructive competition.[1] Underlying such results is the fact that learning from experience directly or indirectly gives rise to comparisons of relative payoffs, so the criterion for evaluation of choices is different from that of payoff maximization. And, for a large class of environments, there is a natural competition between agents: the expansion of activity by one tends to depress the payoffs of others in relative terms so that the new activity level is vindicated by comparison. In this way, a slow incremental deterioration in welfare can go undetected, but ultimately lead to a large drop in welfare, as imitative behavior promotes excessive activity, depressing the welfare of all. In short, "keeping up with the Jones' " ends up leading to the worsening of the Jones' welfare.

These observations raise a significant concern with the use of imitation as a behavioral dynamic; agents should come to realize that "upward matching" of the actions of others is counter-productive. However, when individuals cannot recall and evaluate the performance of past actions, the process can go unnoticed, with payoffs declining, period by period. In contrast, when they have sufficient memory, they can observe the effect of destructive imitation through the evaluation of payoff averages over some period of time. This crucial feature provides an environment where cooperative behavior is sustainable. In essence, individuals have "long enough" to evaluate the performance of different actions. We develop these issues in an environment where agents recall a history of actions of finite length and the payoffs associated with these actions. Agents imitate successful past behavior and this generates dynamics on a state space of finite histories leading naturally to the study of the long-run behavior of the process.

In this long-run setting, it seems plausible that the actions agents take should somehow be justified

[1] See Matros and Josephson (2004), Vega-Redondo (1997) and Alós-Ferrer (2004).

over time and with experience. Having agents evaluate an action through its historical average performance over time does precisely this — they may attach a degree of confidence in the reliability of expectations concerning the value of a given action. Our analysis shows that this confidence is justified and experience-based judgment turns out to be supported by fact: choices, which agents through experience come to believe are good, *do* yield high payoffs. When long histories are observed, agents see the destructive consequences of "ratchet" imitation — in contrast to when memory is short so that trends cannot be detected. Not only does this longer view moderate destructive competition, but imitation can also form the basis for cooperative or collusive behavior. In particular, for a broad class of environments, when recall is sufficiently good, the unique stochastically stable outcome with imitation is the maximally collusive outcome (the monopolistic outcome in oligopoly contexts).

There is ample psychological evidence that imitative behavior of the sort that we model is common (see, e.g., Kahneman et al. (1997)'s review of experimental work on how recalled utility affects individual action choices). Also, this model of imitative behavior has a substantial degree of internal consistency. In the long run, imitating the success of others turns out to produce success for the imitator. The individual follows a pattern of behavior—imitation—and the evidence supports that behavior. Furthermore, although imitative behavior is myopic in the sense that agents do not try to predict the behavior of others and optimize against the prediction, a "sophisticated" agent may be unable to exploit the myopic behavior of others. Although a sophisticated player may be able to increase own current payoff by changing action, this would cause others to imitate the successful behavior, and continue to imitate until the payoff performance of the new choice dropped below the initial status quo, so that the sophisticated player's gain is completely eroded.

The classes of environment considered here are those in which there is strategic covariation of payoffs and strategies. With complementarities, higher actions of another player raise one's payoff and best response; and with substitutes, higher actions of another player lower one's payoff and best response. We use the tools of stochastic stability to study the dynamic evolution of behavior. Few assumptions are imposed on the structure of the stochastic components of the model because, in this framework, virtually any model of experimentation used to identify stochastically stable states leads to the same general conclusion of long-run cooperative behavior. The key insight in the relevant analysis is that the length of memory plays a crucial role because the "size" of the basins of attraction of different states varies with memory length. In comparing two long-run absorbing states, increasing the length of memory increases the relative size of the basin of attraction of the most "collusive" state. As a result, lengthening memory leads to the stochastically stable states being the most collusive or cooperative ones. And, because it is the length of memory rather than the specific model of experimentation that generates this effect, the model of experimentation is largely irrelevant to the prediction of the long-run steady state.

The next section describes the model. Section 3 formulates imitative behavior and details how error and experimentation are introduced into the framework. Section section 4 analyzes long-run dynamic behavior, and presents the results described above. Section 5 considers variations on the imitation rule

to emphasize the robustness of the general point being made; and section 6 concludes.

2 The Model, Notation and Assumptions

In each period, each of n identical agents chooses an action from a common finite set of z actions, $X = \{x_1, ...x_z\}$, with representative element x. The strategy spaces is ordered so that $x_1 < \ldots < x_z$. Write X^n to denote the n-fold product of X, and denote an action profile by $\underline{x} \in X^n$. For notational emphasis, we sometimes write the action profile $\underline{x} = (x^1, \ldots, x^n)$ as (x^i, x^{-i}) in order to highlight the i^{th} component, where x^i is the choice by agent i and x^{-i} is the $n-1$ vector of choices by the other agents.

The payoff to i at action profile (x^i, x^{-i}) is given by $\pi_i(x^i, x^{-i})$. Payoffs are assumed to be symmetric in the sense that there is a function $\pi : X^n \to R$ such that for any i, we have $\pi_i(x^i, x^{-i}) = \pi(x^i, x^{-i})$. At a symmetric action profile where $x^i = x^j = x$ for all i and j, write $\pi^*(x) = \pi(x, x, \ldots, x)$, where $\pi^* : X \to R$. Payoffs are assumed to be strictly increasing for $x < x_m$ and strictly decreasing for $x > x_m$. That is, x_m is the "most collusive" choice; it is the unique solution to $\max_x \pi^*(x)$. In an oligopoly setting, x_m corresponds to the equal ($\frac{1}{n}$th) share of the monopoly output. Given a vector $\underline{x} = (x^1, x^2, \ldots x^i, \ldots x^n)$, emphasize the impact of i's action on different agents' payoffs by writing $(x^1, x^2, \ldots [x^i]^i, \ldots x^n)$. We consider environments in which actions are either substitutes or complements.

Definition 1 *Actions are substitutes if (a) actions are strategic substitutes—best-response functions are monotone decreasing—and (b) an action increase by one agent lowers the payoff of other agents: if $j \neq i$, $x^j > 0$, and $\tilde{x}^i > x^i$ then $\pi^j(x^1, \ldots, [\tilde{x}^i]^i, \ldots, x^n) < \pi^j(x^1, \ldots, [x^i]^i, \ldots, x^n)$.*

Actions are complements if (a) actions are strategic complements—best-response functions are monotone increasing—and (b) an action increase by one agent raises the payoff of other agents — if $j \neq i$, $x^j > 0$, and $\tilde{x}^i > x^i$ then $\pi^j(x^1, \ldots, [\tilde{x}^i]^i, \ldots, x^n) > \pi^j(x^1, \ldots, [x^i]^i, \ldots, x^n)$.

Many familiar economic models fit into this framework. For example,

- The traditional oligopoly model has preferences $\pi_i(x^i, x^{-i}) = p(\sum_{j=1}^n x^j)x^i - c(x^i)$, where $p(\cdot)$ is the inverse demand and $c(\cdot)$ is a firm's cost function.
- The tragedy of the commons has preferences $V(\sum_j x^j)x^i - v(x^i)$, where x^i is the number of owner i's sheep and $V(\sum x^j)$ is the gain to having another sheep on the commons when there are already $\sum x^j$ sheep there.
- In a team production problem, payoffs are $u(w(Q(\sum x^j))) - v(x^i)$, where $Q(\sum_j x^j)$ is total output, which depends on group effort; $w(Q)$ is the wage, which is a function of output; and $v(x^i)$ is the disutility attached to effort x^i.

Denote the symmetric Nash equilibrium action choice by x_N. We assume that if every player is playing a common choice below the Nash equilibrium strategy, then a player can gain by raising their action; and conversely if everyone is selecting a strategy above the Nash level it pays to reduce one's action:

(i) If $x < x_N$, for $x < \tilde{x}^i \leq x_N$, then $\pi^i(x, x, \ldots, [\tilde{x}^i]^i, x, \ldots, x) > \pi^*(x)$;
(ii) if $x \geq x_N$, for $x_N \leq \tilde{x}^i < x$, then $\pi^i(x, x, \ldots, [\tilde{x}^i]^i, x, \ldots, x) > \pi^*(x)$.

This assumption ensures that there is a unique symmetric Nash equilibrium. At a common choice below x_N, a small increase in agent i's action raises own payoff; while at a common choice above x_N, small reductions in action raise payoffs. To avoid confusion, the paper primarily focusses on the substitutes case. A parallel analysis holds for the complements case. One result, theorem 4 is stated for both cases.

3 Behavior: Experience-Based Choices and Experimentation.

As already described, the model of behavior is one in which individual choices are based on experience: information from past choices and payoff consequences is used to guide current choice. In particular, to implement these experience-based choice rules, agents need not know the detailed structure of the environment in which they operate, nor do they need to understand the motivations of other individuals. We next describe the decision-making process. Following that, we introduce a general model of experimentation and error. Experimentation and error, in combination with imitative-based choice, govern the step-by-step movement of the system and its long-run dynamic behavior.

3.1 Experience-Based Choices

Agents observe the actions taken by other agents and the associated payoffs. At the beginning of each date $t+1$, an agent can recall the actions and payoffs for the past $l+1$ periods, $t, t-1, \ldots, t-l$. Because actions uniquely determine payoffs, we conserve on notation and write the state or recalled history of the economy as a $n \times (l+1)$ vector of choices made by agents in the previous $l+1$ periods. In each period individuals observe the choices made and the rewards received; but they may not know how choices determine payoffs, i.e., they may not know the payoff functions.

At the beginning of date $t+1$, the state is given by $s = \mathbf{x}(t) = (\underline{x}_{t-l}, \ldots, \underline{x}_{t-1}, \underline{x}_t)$, where $\underline{x}_\tau \in X^n$ is an n vector of choices by agents, and $t - l \leq \tau \leq t$. Denote the set of states by S. Given a state $s = \mathbf{x}(t)$, let $X(s) = \{x \in X \mid \exists i,\ t - l \leq \tau \leq t,\ x = x^i_\tau\}$, so that $x' \in X(s)$ if at state s some player in the history identified by s chose x'.

At date t, given the state $\mathbf{x}(t)$, considering the average historical performance of each choice, denote the average payoff from the choice x at state s by $\bar{\pi}[x : s]$, calculated by averaging over those periods in which x was chosen and can be recalled. Let $T_t(x \mid \mathbf{x}(t)) = \{\tau \mid t - l \leq \tau \leq t,\ \exists j,\ x^j_\tau = x\}$ be the set of times at which some agent chose x. At each $t' \in T_t(x \mid \mathbf{x}(t))$, let $k(x, t')$ be an arbitrary agent who

played x at that date (symmetry across agents implies that in any period where agents made the same choice, they received the same payoff). The payoff to those agents, including $k(x,t')$, who chose x at time t' is then $\pi^{k(x,t')}(x, x_t^{-k(x,t')})$. The average payoff from action x over the memory frame is therefore

$$\bar{\pi}[x:s] = \frac{1}{\#T_t(x \mid \mathbf{x}(t))} \sum_{t' \in T_t(x|\mathbf{x}(t))} \pi^{k(x,t')}(x, x_t^{-k(x,t')}).$$

If both x and x' are in $X(s)$, then both are in the memory frame and may be compared by all agents. Because agents who take the same action in a period receive the same payoff, results are unchanged if the average is taken over agents as well as time periods.

Agents select choices based on average historical performance: in a stationary environment, it is reasonable to suppose that lessons learned from one period are no more or less valuable than those learned in other periods, in which case experiences from different periods should be weighted equally. In fact, it is important for the results that more recent payoffs from an action be weighted at least as heavily as more distant experiences (so that the effects of imitation are quickly recognized). Formalizing previous discussion, agents select actions from the set $B(\mathbf{x}(t))$ of choices that yielded the highest average historical performance, where

$$B(\mathbf{x}(t)) = \{x \in X(\mathbf{x}(t)) \mid \bar{\pi}[x; \mathbf{x}(t)] = \max_{x' \in X(\mathbf{x}(t))} \bar{\pi}[x'; \mathbf{x}(t)]\}.$$

Represent agent i's action choice by a distribution $b_i(x \mid \mathbf{x}(t))$ on the set of actions, X. The next definition formalizes the idea that agent i makes choices based on comparison of actions from the set of best historical performers.

Definition 2 *Agent i optimizes imitatively according to b_i if at every state $s = \mathbf{x}(t)$, $b_i(\cdot \mid \mathbf{x}(t))$, the choice of action for period $t+1$ has support $B(\mathbf{x}(t))$.*

For generic payoffs, $b_i(\cdot \mid \mathbf{x}(t))$ puts probability 1 on a single point—and in such an event, we denote that point by $b_i(\mathbf{x}(t)) \in X$.

The next section describes the way in which error is added to the choice procedure. This model of error or experimentation in conjunction with (imitative) optimizing behavior determine the dynamic evolution of the system.

3.2 A General Model of Error and Experimentation.

At any state of the system each individual recalls the history of actions $\mathbf{x}(\mathbf{t})$ and corresponding payoff profiles, and with this information has an intended choice, x_i. For all individuals, the intended action profile is $\underline{x} = (x_1, x_2, \ldots, x_n)$. We postulate a general model evolution and experimentation in which exploration of alternatives relative to this action, or errors in choice, is formulated as a distribution φ that selects

the intended action with high probability, but also gives some weight to the selection of other actions.

Definition 3 *A mapping φ_θ is a model of experimentation and/or error if for each $(\underline{x}, \mathbf{x}(t))$, φ_θ is a distribution (parameterized by θ) on $X^n = X \times X \cdots \times X$, (n times) satisfying:*

1. *For all $\underline{\tilde{x}} \in X^n$, $\varphi_\theta(\underline{\tilde{x}} \mid \underline{x}, \mathbf{x}(t)) > 0$.*
2. *$\varphi_\theta(\underline{x} \mid \underline{x}, \mathbf{x}(t)) \approx 1$ and $\varphi_\theta(\underline{x} \mid \underline{x}, \mathbf{x}(t)) \to 1$ as $\theta \to 0$.*

In words, at $(\underline{x}, \mathbf{x}(t))$, every point in X^n has positive probability, but most of the probability mass is placed on the point $\underline{x}$. Experimentation or error is represented by the distribution φ_θ. Absence of such experimentation corresponds to deterministic selection of the intended choice by each individual, which may be represented by the distribution φ^*, where $\varphi^*(\underline{x} \mid \underline{x}, \mathbf{x}(t)) = 1, \forall \underline{x} \in X^n$. Thus, φ_θ is approximately equal to φ^* and $\varphi_\theta \to \varphi^*$ as $\theta \to 0$. Because X is finite, we have $\underline{\varphi}_\theta \equiv \min_{\tilde{x}, x, \mathbf{x}(t)} \varphi_\theta(\underline{\tilde{x}} \mid \underline{x}, \mathbf{x}(t)) > 0$, and $\overline{\varphi}_\theta \equiv \max_{\tilde{x}, x, \mathbf{x}(t)} \varphi_\theta(\underline{\tilde{x}} \mid \underline{x}, \mathbf{x}(t)) \geq \underline{\varphi}_\theta$. Because mutation or experimentation is infrequent, $\overline{\varphi}_\theta$ is small. For reasons discussed in Bergin and Lipman (1996), it is assumed that there is a positive integer k such that $\frac{\overline{\varphi}_\theta^k}{\underline{\varphi}_\theta}$ is bounded by some constant, for θ small: the rates of error or experimentation in different states cannot differ by an infinite number of orders of magnitude.

This formulation imposes few assumptions on experimentation and error. Experimentation may depend on the state (history) $\mathbf{x}(t)$, and may be correlated across individuals. For example, suppose that experimentation depends on the state, s, own choice prior to experimentation, x_i, and some publicly observed random variable, ω with distribution ν: $\varphi_\theta^i(\tilde{x}_i \mid \omega, x_i, s)$. Then the distribution over choices is given by

$$\varphi_\theta(\underline{\tilde{x}} \mid \underline{x}, s) \stackrel{\text{def}}{=} \int_\Omega \times_{i=1}^n \varphi_\theta^i(\tilde{x}_i \mid \omega, x_i, s) \nu(d\omega).$$

With independent experimentation across individuals, φ_θ has the form: $\varphi_\theta(\underline{\tilde{x}} \mid \underline{x}, \mathbf{x}(t)) = \times_{i=1}^n \varphi_\theta^i(\tilde{x}_i \mid x_i, \mathbf{x}(t))$, where $\underline{x} = (x_1, x_2, \ldots, x_n)$ and $\underline{\tilde{x}} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)$.

Further, when for each player, a common choice $x_i \in X$ is the intended action, we can represent state-independent uniform experimentation over alternatives with small probability θ using the multinomial distribution. For example, given the current state, suppose that each individual plans to choose the same action, x_i, next period. In this case, player i chooses x_i next period with probability $(1 - \hat{z}\theta)$, and x_j with probability θ, $j \neq i$, where $\hat{z} = z - 1$ (recalling that z is the number of possible choices or elements of X), so that

$$\text{Prob}(n_j \text{ players choose } x_j, j = 1, \ldots, z) = \frac{n}{\Pi_{i=1}^z n_i!} (1 - z\theta)^{n_i} \Pi_{\substack{i=1 \\ j \neq i}}^z \theta^{n_j},$$

where $\sum_{k=1}^z n_i = n$.

The system evolves from period to period in two steps. First, choices are made, and then a random component impacts decisions (with small probability), to produce next period's state.

3.3 Dynamics under Selection and Experimentation

Combining $\{b_i\}_{i=1}^n$ and φ_θ determines a Markov process on the state space. If $\underline{x} = (x_1, \ldots, x_n)$, let $b(\underline{x} \mid \mathbf{x}(t)) = \times_{i=1}^n b_i(x_i \mid \mathbf{x}(t))$. With experimentation, the distribution over X is given by:

$$p_\theta(x' \mid \mathbf{x}(t)) = \sum_{\underline{x} \in X^n} \varphi_\theta(x' \mid \underline{x}, \mathbf{x}(t)) b(\underline{x} \mid \mathbf{x}(t)), \ \forall x' \in X^n.$$

Letting $\mathbf{x}(t) = (x_{t-l}, x_{t-l+1}, \ldots, x_t)$, the state $\mathbf{x}(t+1) = (x_{t-l+1}, \ldots, x_t, x')$ has probability $p_\theta(x' \mid \mathbf{x}(t))$, so that the transition probability on the state space is $P_\theta(\mathbf{x}(t+1) \mid \mathbf{x}(t)) = p_\theta(x' \mid \mathbf{x}(t))$. Because the Markov chain is irreducible, it has a unique invariant distribution, which we denote by μ_θ.

With dynamics determined by this process, long-run behavior is considered in terms of the proportion of time spent in each state. The transition matrix is $P_\theta = \{P_\theta(\tilde{s} \mid s)\}_{\tilde{s}, s \in S}$. In the long run, iterates of the transition matrix P_θ converge ($P_\theta^t \to P_\theta^*$ as $t \to \infty$) with a corresponding limiting invariant distribution on states, μ_θ: $\mu_\theta P_\theta = \mu_\theta$. The long-run probability of state i is $\mu_\theta(\{s_i\})$. We denote the state at time t by $s(t)$ and index the set of times at which players move by $T = \{0, 1, \ldots\}$. Define $T_i = \min\{t \in \{1, 2, \ldots\} \mid s(t) = s_i\}$ and $T_i = \infty$ if $s(t) \neq s_i$, $\forall t \geq 1$: for any realization of $\{s(t)\}_{t=0}^\infty$, $T_i \geq 1$ gives the smallest positive integer (if one exists) at which the state $s(i)$ is reached. Conditional on $s(0) = s_i$, T_i gives the first return time to state s_i; and $E\{T_i \mid s(0) = s_i\}$ is the expected return time to state s_i, and is related to $\mu_\theta(\{s_i\})$ according to the formula $\mu_\theta(\{s_i\}) = \frac{1}{E\{T_i \mid s(0) = s_i\}}$. Thus, if $\mu_\theta(\{s_i\}) \approx 1$, then with high probability the system stays there, so that $E\{T_i \mid s(0) = s_i\} \approx 1$.

4 Long-Run Outcomes

The following discussion focuses on the long-run behavior of the system from two perspectives. First section 4.1 shows that that when the level of perturbation of the system through error or experimentation is small, in the long run, the system spends most of the time in the collusive state s^*. Then, in section 4.2 taking the rate of error or experimentation as fixed, we derive the impact of lengthening memory. Specifically, we prove that lengthening memory raises the average frequency of time spent in the collusive state — longer memory promotes greater cooperation. With a fixed experimentation rate, the fraction of the time spent in non-collusive absorbing states goes to 0 as memory length increases.

4.1 The Stochastically Stable State

Stochastic stability concerns the behavior of the distribution μ_θ as $\theta \to 0$. The concept has been utilized extensively in economics since its use by Kandori, Mailath and Rob (1993) and Young (1993). Assuming μ_θ converges, say to μ, a state $s \in S$ is called stochastically stable if it is in the support of the limit, μ. In the specific case where the support of the limiting distribution contains a single point or state, this

implies that in the long run, in the perturbed system most of the time the system is in that state, and so it is the appropriate long-run prediction. In what follows, assume that payoffs are generic, so that in computing finite averages, there are no payoff ties comparing different actions.[2]

Theorem 1 *For generic payoffs, if the length of memory is sufficiently large, then the unique stochastically stable state is* $s^* = (\underline{x}^m, \ldots, \underline{x}^m)$.

Proof: All proofs are in the appendix.

The exact structure of the error or experimentation process matters little for this result. What is central to this result is (1) that the stochastic stability criterion introduces perturbations in behavior which leads to comparison of alternative choices, and (2) the comparisons following experimentation are based on the average performance of different actions. Because the collusive choice has the highest average, once this is played for a sufficiently long period of time, it becomes more difficult to "dislodge" as the preferred choice. Apart from possible gain on introduction, new alternatives perform less well than the collusive outcome, averaged over long enough a period of time. For the same reason, multiple innovations in choice apart from the collusive choice average down over time, and again, if memory is long enough, are discarded in favor of a return to the collusive choice. Furthermore, the longer is memory, the more pronounced is this fact. The contrast between this predicted collusive outcome and the standard prediction of the Walrasian outcome is sharp. What drives the latter prediction (see Bergin and Bernhardt (2004)) is that comparisons are based on the *best* recalled historical performance of an action: individuals never revisit the value of an action, if subsequently that action performed less well. But as Kahneman et al. (1997)' experimental evidence makes clear, individuals do recall and weight more recent experiences in addition to peak performances. Effectively, such weighting underlies this result, and the conclusion highlights how this central prediction is, in fact, reinforced when more recent experiences from an action are weighted more heavily in decision making.

The following simple duopoly example illustrates the impact of comparisons based on average performance. We explicitly calculate how long memory has to be to support the monopoly outcome. The example highlights that memory recall does not have to be very long to support collusive behavior.

[2]Consider the action $x \in X$. Let $\{x_\alpha^{-i}\}$ be an enumeration of points in X^{n-1}. For a history of length r, the finite set of possible average payoffs associated with x is

$$\overline{\Pi}(x) = \{u \in R \mid u = \frac{1}{\sum k_\alpha} \sum_{\alpha=1}^{z^{n-1}} k_\alpha \pi(x, x_\alpha^{-i}),\ k_\alpha \text{ integer}, \sum k_\alpha \leq r\}.$$

A Duopoly Example. Consider a duopoly model with three choices — M, N or W, where M denotes the most collusive behavior, N the Nash equilibrium outcome, and W the most competitive behavior.

$$\begin{array}{c} \\ M \\ N \\ W \end{array} \begin{array}{c} \begin{array}{ccc} M & N & W \end{array} \\ \begin{pmatrix} (6,6) & (3,7) & (2,4) \\ (7,3) & (4,4) & (1,2) \\ (4,2) & (2,1) & (0,0) \end{pmatrix} \end{array}$$

To discuss the long-run dynamics, suppose that experimentation is independent with each agent experimenting independently with probability ϵ in each time period. Thus, the probability that both experiment in any period is ϵ^2, which is approximately the probability that exactly one individual experiments in each of any two periods, $(2(1-\epsilon)\epsilon)^2$.

With just two players, a history consists of $l+1$ pairs, $[(c^1_{t-l}, c^2_{t-l}), \ldots, (c^1_t, c^2_t)]$, where c^i_τ is the choice of i in period t. Under imitative dynamics only monomorphic states (where agents take the same action over time) can be stable, so it is sufficient to restrict attention to such states. Consider a history of the form $[(M,M), \ldots, (M,M), (M,M)]$ after which experimentation switches player 2's choice to N — so that next period's history is $[(M,M), \ldots, (M,M), (M,N)]$. The average payoff to M is $\frac{l}{l+1}6 + \frac{1}{l+1}3$, and the average payoff to N is 7. Imitation leads player 1 to switch to N, leading to the history, $[(M,M), \ldots, (M,M), (M,N), (N,N)]$. At this point, with l observations on M the average payoff to M is $\frac{l-1}{l}6 + \frac{1}{l}3$ and the average payoff to N is $\frac{1}{2}7 + \frac{1}{2}4 = \frac{11}{2}$. Hence, the average performance of M is better than N if $\frac{l-1}{l}6 + \frac{1}{l}3 > \frac{11}{2}$, which holds if $l > 6$. If two players experiment simultaneously, yielding history $[(M,M), \ldots, (M,M), (N,N)]$, then reversion to M is immediate. If experimentation occurs twice, one after the other, then history $[(M,M), \ldots, (M,M), (M,N), (N,M)]$ can arise (where M and N are matched twice), and without further experimentation both now play N. One period later, the average payoff to M is $\frac{l-2}{l}6 + \frac{2}{l}3$, whereas the average payoff to N is 7, and both players choose N. Two periods later, the history becomes $[(M,M), \ldots, (M,M), (M,N), (N,M), (N,N), (N,N)]$. The average payoff from M is $\frac{l-3}{l-1}6 + \frac{2}{l-1}3 = \frac{6l-12}{l-1}$, and the payoff from N is $\frac{1}{4}(7+7+4+4) = \frac{11}{2}$. Reversion by both players to l occurs if $\frac{14l-10}{l-1} > \frac{11}{2}$, which holds if $l > 13$.

Now consider the constant Nash history, $[(N,N), \ldots, (N,N), (N,N)]$. Experimentation to M by one player leads to $[(N,N), \ldots, (N,N), (N,M)]$ followed by reversion to $[(N,N), \ldots, (N,N), (N,N)]$, as M produces a lower average payoff than N, so that reversion to N occurs. However, if both experiment with M, the history becomes $[(N,N), \ldots, (N,N), (M,M)]$, and because the average from M is higher, both play M thereafter. Similar calculations apply when the choice of W occurs. Thus, if agents can recall at least 14 periods, the unique stochastically stable outcome has both players choose M.

Note that in the case where players have just two strategies, $\{M, N\}$, the game is a Prisoner's dilemma game and the same reasoning gives $\{M, M\}$, the cooperative outcome, as the long-run outcome.

4.1.1 Long-Run Outcomes: Local Experimentation

In this framework, one can consider local experimentation where only actions in a neighborhood of a choice are tested. Let $d(x^i)$ be the largest action less than x^i and $u(x^i)$ be the smallest action larger than x^i, according to the ordering on X. Given $\underline{x} \in X^n$, define $\mathcal{N}(\underline{x}) = \{\underline{\tilde{x}} \mid \tilde{x}^i \in \{d(x^i), x^i, u(x^i)\}\}$, Experimentation is local if it involves moving up or down to an adjacent action, so that $\varphi_\theta(\cdot \mid x, \mathbf{x}(t))$ has support $\mathcal{N}(x)$. We next prove that argument underlying Theorem 1 holds when experimentation is local.

Theorem 2 *Suppose that experimentation is local and the memory length is sufficiently long. Then the unique stochastically stable state is* $s^* = s_{x_m} = (\underline{x}^m, \ldots, \underline{x}^m)$.

The next section considers the impact of lengthening memory. The analysis is complicated by the fact that increasing memory length changes the state space. For this reason, it is more convenient to focus on basins of attraction, because with long memory, a state may involve long "strings" of collusion or cooperation, without being the (unique) cooperative state.

4.2 The Impact of Lengthening Memory

With memory length l, a state of the system at time t is given by $s = \mathbf{x}(t) = (\underline{x}_{t-l}, \underline{x}_{t-1}, \underline{x}_t)$, a vector of the current and l previous periods of the history. The state of the system associated with collusion is $s^* = (\underline{x}_m, \ldots, \underline{x}_m)$, where $\underline{x}_m$ is repeated $l+1$ times. For fixed l, sufficiently large, our previous analysis revealed that when for each $\underline{x}$, the distribution $\varphi_\theta(\cdot \mid \underline{x}, \mathbf{x}(t))$ puts most of the weight on $\underline{x}$, then the state s^* is occupied a large fraction of the time. The maximum error rate or rate of experimentation is bounded above by $\overline{\varphi}_\theta$: $\forall(\underline{x}, \mathbf{x}(t))$, for $\underline{x}' \neq \underline{x}$, $\varphi_\theta(\underline{x}' \mid \underline{x}, \mathbf{x}(t)) \leq \overline{\varphi}_\theta$, so the average time between experimentation or choice error is at least $1/\overline{\varphi}_\theta$. Under this perturbed system, the invariant distribution is μ_θ and from the earlier discussion, $\mu_\theta(\{s^*\}) \to 1$ as $\theta \to 0$. To emphasize the memory length, write s_l^* instead of s^*, let $Q(s_l^*)$ be the basin of attraction of s_l^* in the unperturbed system with memory length l, and let $Q^c(s_l^*)$ be the complement of $Q(s_l^*)$. Finally, write μ_θ^l for the corresponding invariant distribution of the perturbed system with memory length l and perturbation parameter θ. The following theorem shows that as the length of memory increases, the time spent at the collusive state or in its basin of attraction converges to 1. Because experimentation or error rates are constant over time, over a period of time of $1/\overline{\varphi}_\theta$, on average, experimentation will occur so that within the basin of attraction there is a lower bound on moving from one state to another. Nevertheless:

Theorem 3 *With the rate of experimentation fixed sufficiently small ($\overline{\varphi}_\theta$ a fixed small number), the share of total time spent in the basin of attraction of the collusive outcome converges to 1 as the memory length goes to infinity:*

$$\lim_{l \to \infty} \left\{ \frac{\mu_\theta^l(Q(s_l^*))}{\mu_\theta^l(Q^c(s_l^*))} \right\} \to \infty.$$

When the error or experimentation rate is fixed at some small positive level, the system will move from one state to another over time — it cannot settle down in any state. And, for fixed experimentation rate, as memory becomes longer, experimentation is more likely to occur over the time frame of memory length. But, starting from the collusive monomorphic state, as memory length increases, the probability of leaving the basin of attraction of the collusive state goes to 0. In essence, there is more time to "wash out" experiments and subsequent behavior, then returning to collusive behavior with the higher recalled average payoff. To move the system to a different basin of attraction requires repeated experimentation of sufficient frequency relative to the memory length in order to "purge" memory of the high payoff collusive behavior. In contrast, from a non-collusive monomorphic state, experimentation which achieves the collusive per period action profile is immediately adopted and the the system is in the basin of attraction of the collusive state in one step. Now, additional shocks are required to *prevent* the system from moving to the collusive state. Again, the key point is not the role of the error-experimentation process, but rather that fact that long memory makes it more difficult to dislodge high payoff collusive behavior from memory, and once reached, it becomes the benchmark for judging other choices.

5 Other Learning Rules.

Throughout, our analysis has considered the "learning rules" rule of selecting the best historical performer and we have shown that it leads to collusive behavior. At the other end of the spectrum, one can consider the rule of avoiding the worst average historical performer. We next make the central point that even with this very "weak" improving criterion of avoiding worst performing actions, there is still considerable drift to cooperative behavior in the long run. To simplify the discussion, we focus on experimentation that is independent across individuals and independent of the state, using the formulation mentioned in section 3.2, where an individual chooses the intended strategy with probability $(1-\hat{z}\theta)$, and each alternative with probability θ.

A minimal requirement for reasonable learned behavior is that agents not choose the worst historical outcome whenever there is a better alternative. The worst-performing choice solves: $\min_{x' \in X(\mathbf{x}(t))} \bar{\pi}[x'; x(t)]$. If not all choices yield the lowest payoff, let

$$S(\mathbf{x}(t)) = \{x \in X(\mathbf{x}(t)) | \bar{\pi}(x; \mathbf{x}(t)] > \min_{x' \in X(\mathbf{x}(t))} \bar{\pi}[x'; \mathbf{x}(t)]\}$$

be the set of actions with average payoffs that strictly exceed the minimum. If the average payoff to every action chosen is the same (because every action that can be recalled yielded the same historical average performance), then let $S(\mathbf{x}(t))$ correspond to the set of actions that have been played in the last $l+1$ periods: $S(\mathbf{x}(t)) = \{x \in X(\mathbf{x}(t))\}$. Write $\mathcal{S}$ for the class of S's that satisfy these conditions. A minimal restriction on agent behavior is that an agent not select a worst performer, when that is possible:

Definition 4 *Agent i avoids worst choices according to imitation criterion $S \in \mathcal{S}$ if at every state $s =$*

$\mathbf{x}(t)$, *his action choice for period* $t+1$ *is drawn from a probability distribution* $\gamma^i(s)$ *with support* $S(\mathbf{x}(t))$.

As before, the error-experimentation process is applied to this selection procedure. The key point of the next theorem is that, even with this minimal requirement, substantial cooperation arises in the long run. In the statement of Theorem 4 x_{m+1} is the smallest action choice exceeding the collusive action level x_m, and X_{N+1} is the smallest choice exceeding the Nash action level x_N.

Theorem 4 *Suppose that* l *is large, and agents optimize imitatively according to some* $S \in \mathcal{S}$. *Then if actions are substitutes, the set of stochastically stable states is a subset of* $\{x_m, x_{m+1}, \ldots, x_N\}$. *If actions are complements the set of stochastically stable states is a subset of* $\{x_N, x_{N+1}, \ldots, x_m\}$.

To see what underlies this result, consider a homogeneous oligopoly and suppose that the system is at rest at some output level between Cournot and Walrasian. Now consider the impact of perturbing the output of some firm. First consider a small increase in the output by firm i. Relative to the initial quantity, i's profit is below the status quo profit, because the status quo output exceeded Cournot. In the period following the perturbation, i's profit exceeds that of other firms whose profit was depressed by the increase in i's output. But when this lower profit of the other firms is averaged over all periods in which the status quo output was chosen, the single period has a small impact on the average. In contrast, i's profit has fallen relative to the average. Hence, this upward perturbation in output is not adopted. Now suppose i's output drops a small amount. Then, i's profit rises, and the profit of the other firms from playing the status quo, averaged over those periods in which the status quo was played, is approximately unchanged. Hence, the average profit from the lower output exceeds the profit from the status quo output. As a result, all firms adopt this lower output level. This further raises the profit obtained from output reduction, so this lower output persists.

6 Conclusion

The results in this paper show that long memory in conjunction with imitative behavior supports maximally cooperative or collusive behavior. But even when recall is short, so that the scope for a 'good" (cooperative) action to be recognized as good in terms of payoffs is limited, some cooperative behavior can arise as long as memory is non-trivial. To see this, consider the oligopoly setting with the shortest non-trivial memory of two periods, and suppose that the output choice grid is sufficiently fine. Then it can be shown that the stochastically stable outputs are strictly between monopoly and the Walrasian or fully competitive output with price equal to marginal cost. Suppose that the system is at rest at the Walrasian output. But then a slight reduction in one firm's output raises the profit of firms producing the Walrasian output relative to the experimenting firm, but by a factor of *less than two*. Hence, when the payoff from the Walrasian output is averaged over time, it is less than the profit accruing to the experimenting firm. Conversely, monopoly cannot be sustained because it takes more than two periods

to learn that a slight upward experimentation is unprofitable: at x_m, $\frac{\partial \pi^i}{\partial x^i} \mid_{\underline{x}_m} = -(n-1)\frac{\partial \pi^j}{\partial x^i} \mid_{\underline{x}_m} > 0$, where n is the number of firms.

The analysis has assumed that agents weight experiences in different periods from the same action equally. This makes sense in a stationary environment. Still, it may be reasonable to suppose that payoffs from more recent actions are weighted more heavily, because agents suspect that the environment broadly interpreted as reflecting the actions of other agents may have changed. To make this transparent, consider an oligopoly setting and suppose that only the *most* recent experience from an action is weighted. A moment's reflection reveals that it now only takes a memory of three periods to support the monopoly outcome. To see this, suppose that the system is at the collusive outcome, and someone profitably experiments with increased output. Then, in the next period, everyone imitates this increased output, reducing everyone's profit below the monopoly levels. Since firms now only weight recent experiences from an action, they immediately return to the monopoly output level, which they still recall. Conversely, suppose that the system is at rest at outputs above monopoly, and all firms experiment with monopoly. Then it is adopted and retained. Thus, weighting recent payoffs from actions more heavily is *very different* from having a trivial memory.

Finally, we have supposed that individuals select those actions that yielded the best average historical performance. Over time, the consequence of this is to generate collusive behavior. Consider the alternative model of behavior in which some agents optimize explicitly. Although except in special cases, it is difficult to examine the dynamics when one player best responds against the imitative rule, it is worth noting in the substitutes model that the effect of best responding in the state s^* is to raise the choice level of others, as they imitate the best response choice. As a result, the gain from best responding is eroded subsequently as other players imitate. In turn, it can sometimes be optimal for a "rational player" to act as an imitator, when others are also doing so.

7 Appendix

Theorem 1 *For generic payoffs, if the length of memory is sufficiently large, then the unique stochastically stable state is* $s^* = (\underline{x}^m, \ldots, \underline{x}^m)$.

Proof: Call an action profile $x = (x^1, \ldots, x^n)$ *monomorphic* if $x^i = x^j$ for all i, j. A state $s = (\underline{x}_{t-l}, \ldots, \underline{x}_t)$ is a *monomorphic state* if there is a monomorphic profile $\underline{x}$ such that $\underline{x}_t = \underline{x}_{t-j}$, $j = 0, \ldots, l$. For $x \in X$, write s_x to denote the monomorphic state associated with x. Under the unperturbed dynamic, φ^*, all non-monomorphic states are transient and every monomorphic state is absorbing: if the system is at a monomorphic state, then it stays in that state thereafter. Thus, if $\hat{\mu}$ is an invariant distribution of P_{φ^*}, then it has support on the set of monomorphic states.

Now, consider the dynamics under φ_θ. Suppose that $x \neq x_m$ and let s_x be the associated monomor-

phic state. Under φ_θ experimentation occurs with small probability. Let $\underline{x}^m = (x_m, \ldots, x_m)$ and $\underline{x} = (x, \ldots, x)$, and by assumption, we have $\varphi_\theta(\underline{x} \mid \underline{x}, s_x) \approx 1$ and $\varphi_\theta(\underline{x}^m \mid \underline{x}, s_x) \geq \underline{\varphi}_\theta$. But, because $x \neq x^m$, we have $\pi^*(x) < \pi^*(x^m)$, so that with probability of at least $\underline{\varphi}_\theta$, all players switch to x_m: in the absence of further error or experimentation the system moves in l periods to state $s^* = (\underline{x}^m, \ldots, \underline{x}^m)$.

Next, consider the situation where the system is in state $s^* = (\underline{x}^m, \ldots, \underline{x}^m)$. Suppose now that the system is perturbed so that some $\underline{\tilde{x}} \neq (x, \ldots, x)$ is drawn (according to the distribution $\varphi_\theta(\underline{\tilde{x}} \mid \underline{x}_m, s^*)$). Next period the history or state has the form $s' = (\underline{x}^m, \ldots, \underline{x}^m, \underline{\tilde{x}})$. Comparing points in $\tilde{x} = (\tilde{x}_1, \ldots, \tilde{x}_n)$, suppose without loss of generality that the choice by player j of $x_j = \hat{x}$ gave the highest payoff, and write $\tilde{x} = (\hat{x}, \tilde{x}_{-j}$. If $\pi(\hat{x}, \tilde{x}_{-j}) < \pi^*(x_m)$, then reversion to x_m is immediate, because the average payoffs of actions other than x_m are based on the one period of observation. If, instead, $\pi(\hat{x}, \tilde{x}_{-j}) > \pi^*(x_m)$, then all players switch to $\hat{x}$ next period, so that the payoff to each player is $\pi^*(\hat{x})$, and the average payoff to $\hat{x}$ is $\frac{1}{2}[\pi(\hat{x}, \tilde{x}_{-j}) + \pi^*(\hat{x})]$. If $\hat{x}$ is chosen for q subsequent periods, then the average payoff to $\hat{x}$ is $\frac{1}{q}\pi(\hat{x}, \tilde{x}_{-j}) + \frac{q-1}{q}\pi^*(\hat{x})$. If $\hat{x}$ was the only choice in the list $\tilde{x}$ with payoff above $\pi^*(x_m)$, then compare this with the payoff from x_m. Depending on whether x_m was in the list $\tilde{x}$, the average payoff to x_m is either $\pi^*(x_m)$ or $\frac{1}{l+1}\pi(x_m, \tilde{x}_{-i}) + \frac{l}{l+1}\pi^*(x_m)$. Because $\pi^*(\hat{x}) < \pi^*(x_m)$, and l is (sufficiently) large relative to q, $\frac{1}{q}\pi(\hat{x}, \tilde{x}_{-j}) + \frac{q-1}{q}\pi^*(\hat{x}) < \min\{\pi^*(x_m), \frac{1}{l+1}\pi(x_m, \tilde{x}_{-i}) + \frac{l}{l+1}\pi^*(x_m)\}$. Taking q_1 to be the smallest integer such that this inequality holds, reversion to x_m occurs after q_1 periods. If there was more that one point in $\tilde{x}$ with payoff above $\pi^*(x_m)$ they are played in order, moving down successively from one to the next as the average payoffs fall: each can be played only for a fixed number of times before the average payoff is below $\pi^*(x_m)$. With l sufficiently large reversion to $\pi(x_m)$ eventually occurs.

Suppose that experimentation occurs at multiple time periods, $\tilde{x}(1), \ldots \tilde{x}(k)$ at $t_0 < t_1 < \cdots < t_k$ where at t_0 the (initial) state is s^*. If l is large relative to k, the historical average payoff to x_m has the form $\frac{k}{l}\sum_{r=1}^{k}\pi_r + \frac{l-k}{l}\pi^*(x_m) \approx \pi^*(x_m)$, when l is large and x_m is matched with actions other than x_m on k occasions.

From the previous paragraph, there is a fixed length of time following the draw of $\tilde{x}(0)$, before reversion to s^* occurs, absent further perturbations. If further perturbations occur sufficiently quickly, t_1 "close" to t_0, other actions may be played for a period of time until averages drop below competing choices that arose from the perturbation, or else below $\pi(x_m)$. If $t_{i+1} - t_i$ is larger than some number, q_{i+1} say, reversion to x_m occurs between t_i and t_{i+1}, so that players are playing x_m when the perturbation associated at t_{i+1} occurs. In this case, the discussion proceeds as above. To prevent reversion to x_m in the time interval from t_0 to t_k it must be that $q = \sum_{r=1}^{k} q_k > t_k - t_0$. But if l is sufficiently large relative to q, if there are no further perturbations after the k^{th}, then reversion to x_m occurs. For any k, there is an l sufficiently large such that k perturbations in any order disturb the system, but it subsequently reverts to x_m in the state s^*.

Thus, for any k, there is an l such that k perturbations will not move the system from state s^*: any k perturbations leave the system in the basin of attraction of the state s^* in the unperturbed system φ^*.

Note that because $\underline{\varphi}_\theta$ and $\overline{\varphi}_\theta$ are small positive numbers, for some integer k, $\overline{\varphi}_\theta^k < \underline{\varphi}_\theta$. If k is chosen to satisfy $\overline{\varphi}_\theta^k < \underline{\varphi}_\theta^2$, then the probability of escaping the basin of attraction of state s^* is an $(\underline{\varphi}_\theta)$ order of magnitude smaller than that of moving from any monomorphic state to s^*. ∎

Theorem 2 *Suppose that experimentation is local and the memory length is sufficiently long. Then the unique stochastically stable state is* $s^* = s_{x_m} = (\underline{x}^m, \ldots, \underline{x}^m)$.

Proof: The proof follows directly from the proof of theorem 1. Specifically, given $x \in X$, $x > x_m$, $\underline{x} = (x, \ldots, x) \in X^n$ and $\underline{x}' = (d(x), \ldots, d(x)) \in X^n$ is reached with probability $\underline{\varphi}_\theta > 0$. Taking k such that $\overline{\varphi}_\theta^k < \underline{\varphi}_\theta$, the same calculations as above show that the system has a higher probability of moving down from x than up. ∎

Theorem 3 *With the rate of experimentation fixed sufficiently small ($\overline{\varphi}_\theta$ a fixed small number), the share of total time spent in the basin of attraction of the collusive outcome converges to 1 as the memory length goes to infinity:*

$$\lim_{l \to \infty} \left\{ \frac{\mu_\theta^l(Q(s_l^*))}{\mu_\theta^l(Q^c(s_l^*))} \right\} \to \infty.$$

Proof: With S the set of states, and $\hat{s} \in S$, an $\hat{s}$-tree is a collection of ordered pairs, $h = \{(s, \tilde{s}) \mid (s, \tilde{s}) \in S \times S\}$ such that for every $s \in S \setminus \hat{s}$ there is a unique successor (some s' such that $(s, s') \in h$); and $\hat{s}$ has no successor. Let $H_{\hat{s}}^l$ be the set of all $\hat{s}$-trees, where the memory length is l. Letting $q_h = \times_{(s,s') \in h} p_{ss'}^\theta$, from the Markov chain tree theorem, the relative probability of s' to s'' in the invariant distribution is given by the ratio: $\frac{\mu(\{s'\})}{\mu(\{s''\})} = \frac{\sum_{h \in H_{s'}^l} q_h}{\sum_{h \in H_{s''}^l} q_h}$. *For ease of notation* in what follows write μ instead of μ_θ^l, always bearing in mind that the invariant distribution μ depends on both l and θ. Similarly, write $\overline{\varphi}$ and $\underline{\varphi}$ instead of $\overline{\varphi}_\theta$ and $\underline{\varphi}_\theta$.

Because there are z monomorphic states, there is an s^*-tree h, such that $q_h \geq O(\underline{\varphi}^{z-1})$. (The notation $O(x)$ denotes "of order of magnitude x" — the ratio of the term to x is bounded.) Consider a memory of length l and state $\hat{s}$ that is absorbing under the unperturbed system. Then the state, $\hat{s}$, is monomorphic so that for some $x \in X$, $\underline{x} = (x, x, \ldots, x)$ and $\hat{s} = s_x = (\underline{x}, \ldots, \underline{x})$; and let $h \in H_{\hat{s}}$. Consider the experimentation or error that produces the movement $\hat{s} \rightsquigarrow (\underline{x}, \ldots, \underline{x}; \underline{x}^m) = \hat{s}^1$. Let $\hat{s}^r$ denote the state with a history of $\underline{x}^m$ in the most recent r periods and $\underline{x}$ in the $l + 1 - r$ periods prior to that. Then, absent further shocks to the system, the system moves deterministically according to the sequence $\hat{s}^1 \to \hat{s}^2 \to \cdots \to \hat{s}^{l+1} = s^*$. Under the structure of error-experimentation, $p_{\hat{s}\hat{s}^1}^\theta$ is bounded below by $\underline{\varphi}$; and for $r = 1, \ldots, l$, $p_{\hat{s}^r \hat{s}^{r+1}}^\theta$ is of order 1: $p_{\hat{s}^r \hat{s}^{r+1}}^\theta = O(1)$ (i.e., $p_{\hat{s}^r \hat{s}^{r+1}}^\theta \approx 1$). Consequently, $\times_{r=0}^l p_{\hat{s}^r \hat{s}^{r+1}}^\theta \geq O(\underline{\varphi})$, taking $\hat{s}^0 = \hat{s}$. The same reasoning applies to each (monomorphic) absorbing state different from s^*. Let $S_M \subset S$ be the set of monomorphic states, and let $S_M^* = S_M \subset S \setminus \{s^*\}$ be the set of monomorphic states, excluding s^*. The set of non-monomorphic states, $S \setminus S_M$ can be partitioned

into (disjoint) sets such that for each state, s, in a given element of the partition there is a collection of distinct states, $\{s_j\}_{j=1}^J$ with $s = s^j$, $s^J = \hat{s}$ and $p_{s^j s^{j+1}} = O(1)$. For each $\hat{s} \in S_m \setminus \{s^*\}$, the chain $\hat{s} = \hat{s}^1 \to \hat{s}^2 \to \cdots \to \hat{s}^{l+1} = s^*$ provides weight of order $\underline{\varphi}$ to the s^*-tree; transient states provide weight of order 1. Because there are (M) monomorphic states, this construction yields an s^*-tree, h, with $q_h \geq O(\underline{\varphi}^{z-1})$.

Next, consider an s-tree for some $s \in S_M^*$. Each of the $z-2$ states in $S_M^* \setminus \{s^*\}$ is absorbing in the unperturbed system. From the previous discussion, for any such state the perturbation $\hat{s} \rightsquigarrow (\underline{x}, \ldots, \underline{x}; \underline{x}^m) = \hat{s}^1$ starts a process that terminates at s^*. The probability of this in the perturbed system has order of magnitude bounded above by $\overline{\varphi}$. Now, consider state s^* and the number of perturbations required to reach the basin of attraction of an alternative monomophic state. Recall from theorem 1 that for any k, there is a memory length $l(k)$ such that k perturbations in any order over l periods do not move the system out of the basin of attraction of s^*. In the s-tree, the chain connecting s^* to s, $\{\tilde{s}^j\}_{j=1}^m$ with $s^* = \tilde{s}^1, \ldots, \tilde{s}^m = s$ has the property that $\times_{j=i}^m p^\theta_{\tilde{s}^j \tilde{s}^{j+1}} \leq O(\overline{\varphi}^k)$, so a bound on the product of probabilities in an s-tree is given by $q_h \leq O(\overline{\varphi}^{z-2+k})$. So, comparing the states s and s^*, $\frac{\mu(\{s\})}{\mu(\{s^*\})} \leq O(\left(\frac{\overline{\varphi}}{\underline{\varphi}}\right)^{z-1} \cdot \overline{\varphi}^{k+1})$. Because $\left(\frac{\overline{\varphi}}{\underline{\varphi}}\right)$ is fixed, as k increases (with corresponding increase in memory length), we have $\frac{\mu(\{s\})}{\mu(\{s^*\})} \to 0$. Furthermore, for any state, s' in the basin of attraction of s, and any state $\tilde{s}$ in the basin of attraction of s^*, $Q(s^*)$ such that more than k perturbations is required to leave $Q(s^*)$, the same inequality holds: $\frac{\mu(\{s'\})}{\mu(\{\tilde{s}\})} \leq \alpha_l \to 0$ as $l \to \infty$.

The state space may be partitioned into a collection of basins of attraction, one for each monomorphic state. Write f_j^l for the fraction of all states associated with monomorphic state j, denoting f_*^l for the fraction associated with state s^*, and f_{k*}^l for the fraction of states in $Q(s^*)$ that require more that k perturbations to leave the basin of attraction $Q(s^*)$ (denoting the set of such states $Q_k(s^*)$). Taking the expression, $\mu(\{s'\}) \leq \alpha_l \, \mu(\{\tilde{s}\})$, summing over $s' \in Q(s_j)$ and $\tilde{s} \in Q_k(s^*)$ gives

$$\left\{ \frac{f_{k*}^l}{f_j^l} \right\} \cdot \mu(Q(s_j)) \leq \alpha_l \mu(Q_k(s^*)) \leq \alpha_l \mu(Q(s^*))$$

Because $\frac{f_{k*}^l}{f_j^l}$ is bounded away from 0 as l increases, and because $\alpha_l \to 0$, $\frac{\mu(Q(s_j))}{\mu(Q(s^*))} \to 0$. Since there are a (fixed) number of monomorphic states and this holds for each one so that $\frac{\mu(Q^c(s^*))}{\mu(Q(s^*))} \to 0$, where recall, $Q^c(s^*)$ is the complement of the basin of attraction of s^*. ■

Theorem 4 *Suppose that l is large, and agents optimize imitatively according to some $S \in \mathcal{S}$. Then if actions are substitutes, the set of stochastically stable states is a subset of $\{x_m, x_{m+1}, \ldots, x_N\}$. If actions are complements the set of stochastically stable states is a subset of $\{x_N, x_{N+1}, \ldots, x_m\}$.*

Proof: Actions are substitutes. Suppose that the system is at rest at state $s = s_x$, where $x > x_N$, and consider a mutation for player i to $x' > x$. (Recall that $\underline{x} = (x, x, \ldots, x)$ and $s_x = (\underline{x}, \ldots, \underline{x})$.)

Agent i earns a payoff of $\pi^i(x,\ldots,x,[x']^i,x\ldots,x)$. Because $x \geq x_c$, we have $\pi^i(x,\ldots,x,[x']^i,x\ldots,x) < \pi^*(x)$, and $\pi^j(x,\ldots,x,[x']^i,x,\ldots,x) < \pi^j(x,\ldots,x,[x]^i,x,\ldots,x)$. Next period, the history or state is s' and consists of all n players playing x for l periods, $n-1$ players playing x in the most recent period and i playing x' in that period. For a player $j \neq i$, the payoff in the most recent period is $\pi^j(x,\ldots,x,[x']^i,x,\ldots,x)$. Because there is just one observation, the average payoff to i from x' is:

$$\bar{\pi}(x' : s') = \pi^*(x) + [\pi^i(x,\ldots,[x']^i,\ldots x) - \pi^*(x)].$$

Now, in all $l+1$ periods x was chosen — yielding a payoff of $\pi^*(x)$ in l periods and $\pi^j(x,\ldots,x,[x']^i,x,\ldots,x)$ in the most recent period. Thus, the average payoff from x at state s' is

$$\begin{aligned}\bar{\pi}(x : s') &= \frac{1}{l+1}[l \cdot \pi^*(x) + 1 \cdot \pi^j(x,\ldots,x,[x']^i,x,\ldots,x)] \\ &= \pi^*(x) + \frac{1}{l+1}[\pi^j(x,\ldots,x,[x']^i,x,\ldots,x) - \pi^*(x)].\end{aligned}$$

Because both $[\pi^i(x,\ldots,[x']^i,\ldots x) - \pi^*(x)]$ and $[\pi^j(x,\ldots,x,[x']^i,x,\ldots,x) - \pi^*(x)]$ are negative, the average payoff from x exceeds the average payoff from x' if:

$$[\pi^i(x,\ldots,[x']^i,\ldots x) - \pi^*(x)] < \frac{1}{l+1}[\pi^j(x,\ldots,x,[x']^i,x,\ldots,x) - \pi^*(x)].$$

For l sufficiently large, this inequality is satisfied. Therefore, all agents revert to choosing x in subsequent periods.

Now consider a downward deviation. In particular, let the deviation x' be the greatest feasible action less than x. Then, $\pi^i(x,\ldots,[x']^i,\ldots,x) > \pi^*(x)$ and $\pi^j(x,\ldots,x,[x']^i,x,\ldots,x) > \pi^*(x)$. From the same calculations, the average payoff from x' exceeds the average payoff from x if:

$$[\pi^i(x,\ldots,[x']^i,\ldots x) - \pi^*(x)] > \frac{1}{l+1}[\pi^j(x,\ldots,x,[x']^i,x,\ldots,x) - \pi^*(x)].$$

Again, for l large, this inequality is satisfied, so that all agents choose x' in subsequent periods. Therefore, if l is large and $x > x_c$, then any upward movement in the action choice by any player is unattractive according to the average payoff performance criteria, but there is a downward movement that is attractive. Posed in an oligopoly context, no first order mutation can raise output, but there exist first order mutations that move the system to a lower common output level by imitation.

Now consider $x < x_m$. Any downward mutation to $x' < x$ by agent i leads to a lower payoff than $\pi^*(x)$, and it raises the payoffs of other agents. Hence, in subsequent periods, agents return to x. In contrast, a small upward mutation to $x' \leq x_m$ by agent i leads to a higher payoff than $\pi^*(x)$, and it lowers the payoffs of other agents. Hence, it is adopted by everyone in the next period. But since $x < x_m$, even when x' is adopted by everyone, each agent's payoff still exceeds that from x. Hence, x'

is adopted.[3]

Actions are complements. The argument is similar. For $x < x_N$, if l is sufficiently large, a small upward mutation by player i to x' yields a greater average payoff than x, and hence is adopted by everyone. In turn, adoption by everyone further raises the payoff from x', so that all agents continue to play x'. In contrast, a downward mutation results in a lower payoff, and (if l is large) is below the average payoff from x, so that all agents eventually return to playing x.

For $x_N < x \leq x_m$, a downward mutation by i to x' may lead to a higher payoff. If it does, then it is adopted by everyone. But then $\pi^*(x') < \pi^*(x)$, so that if l is large, everyone eventually returns to x. An upward mutation by i to $x' > x$ results in a lower payoff to i and a higher payoff to others, so that all agents again play x.

For $x > x_m$, a small downward mutation by i to x' raises his payoff and lowers the payoffs of other agents. Because $x > x_m$, $\pi^i(x, \ldots, x, [x']^i, x, \ldots, x) > \pi^*(x) > \pi^j(x, \ldots, x, [x']^i, x, \ldots, x)$, it is adopted. In turn, because $x > x_m$, $\pi^*(x') > \pi^*(x)$, all agents continue to play x'. Finally, an upward mutation by i to x' lowers his payoffs, and raises everyone else's, so that i returns to x. ■

8 References

[1] Alós-Ferrer, C. , (2004), "Cournot vs. Walras in Dynamic Oligopolies with Memory", *International Journal of Industrial Organization*, Vol. 22, No. 2, February 2004, pp. 193-217.

[2] Bergin, J. and D. Bernhardt (2004), "Comparative Learning Dynamics", *International Economic Review*, 42, No. 2, 431-465.

[3] Bergin, J. and B. Lipman (1996): "Evolution with State-Dependent Mutations", *Econometrica*, 64, 943-956.

[4] 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.

[5] Foster, D. and P. Young, (1990), "Stochastic Evolutionary Games", *Theoretical Evolutionary Biology*, 38, 219-232.

[6] Fudenberg, D. and D. Levine, *The Theory of Learning in Games*, MIT press, 1998.

[7] Kahneman, D., P. Wakker, and R. Sarin, (1997),"Back to Bentham? Explorations of Experienced Utility", *Quarterly Journal of Economics*, 112, 375-405.

[8] Kandori, M., G. Mailath, and R. Rob (1993): "Learning, Mutations and Long-Run Equilibria in

[3] Once the mutation "costs" are determined, the construction of minimum cost trees is straightforward. Consider a choice $x > x_N$ and a tree with root s_x. Pick some $x' < x$, say $x' = x_N$, the Nash level. At some location in the tree, s_{x_N} is placed. Break the edge leaving s^{x_N}, arrange all transient states that are in the basin of attraction of s_{x_N} as predecessors of s_{x_N}. Place s_{x_N} as the root of the tree and form an edge (requiring one mutation) from s_x to a transient state in the basin of attraction of s_{x_N}. This reduces the cost of the tree — so that a tree with root s_x cannot be a minimum cost tree. See Ellison (2000) for discussion of these issues.

Games", *Econometrica*, 61, 29-56.

[9] Matros A. and Jens Josephson (2004), "Stochastic Imitation in Finite Games," *Games and Economic Behavior*, 49, (2004), 244-259.

[10] Vega-Redondo, F. (1997): "The Evolution of Walrasian Behavior", *Econometrica*, 65, 375-384.

[11] Young, P. (1993): "The Evolution of Conventions", *Econometrica*, 61, 57-84.