Dynamic Exploitation of Myopic Best Response*

Burkhard C. Schipper†

October 9, 2018

Abstract

How can a rational player manipulate a myopic best response player in a repeated two-player game? We show that in games with strategic substitutes or strategic complements the optimal control strategy is monotone in the initial action of the opponent, in time periods, and in the discount rate. As an interesting example outside this class of games we present a repeated "textbook-like" Cournot duopoly with non-negative prices and show that the optimal control strategy involves a cycle.

Keywords: Strategic teaching, learning, adaptive heuristics, dynamic optimization, strategic substitutes, strategic complements, myopic players.

JEL-Classifications: C70, C72, C73.

*I thank Rabah Amir, the associate editor, and three anonymous reviewers as well as participants at the 2007 Stony Brook Game Theory Festival for helpful comments. I also thank Zhong Sichen for very able research assistance. This work developed from a joint experimental project with Peter Dürsch, Albert Kolb, and Joerg Oechssler (Duersch, Kolb, Oechssler and Schipper, 2010). Financial support by NSF CCF-1101226 is gratefully acknowledged.

†Department of Economics, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA, Fax: +1-530-752 9382, Email: bcschipper@ucdavis.com

1 Introduction

The question of how to get to equilibrium is arguably as old as the notion of equilibrium. Already Cournot (1838) suggested that firms may reach equilibrium in a quantity setting duopoly by using myopic best response learning. By now there is a large literature on learning in games (for monographs reviewing the literature, see Fudenberg and Levine, 1998, Young, 2013, Hart and Mas-Colell, 2013). This literature focused on "symmetric learning": All players follow the same learning heuristic. Yet, as abilities of players may differ in the real world, it is natural to ask whether a more sophisticated player could strategically teach other learning players and manipulate them to her advantage. Strategic teaching has received only limited attention in the literature on learning (last chapter of Fudenberg and Levine, 1998, Schipper, 2017, Hyndman, Ozbay, Schotter, and Ehrblatt, 2012, Camerer, Ho, and Chong, 2002, Duersch, Kolb, Oechssler, and Schipper, 2010, and Terracol and Vaksmann, 2009). Although some learning heuristics are quite simple (sometimes to the extent that players seem oblivious to the strategic aspects of the interaction), it may still be a non-trivial problem to find the optimal strategy against it. In the case in which the opponent learns according to myopic best response, we are able to characterize the optimal strategy in large relevant classes of two-player games that include games in which a each player's payoff function satisfies strategic substitutes or complements and positive or negative externalities. We think that focusing on myopic best response learning provides a natural starting point because it is arguably the first learning heuristic that has been studied in a game (Cournot, 1838), it has been widely studied in various classes of games (e.g., Dubey, Haimanko, and Zapechelnyuk, 2006, Kukushkin, 2004, Monderer and Shapley, 1996), and many other learning heuristics retain some features of best response learning (e.g., Hart and Mas-Colell, 2006).

For the sake of concreteness, consider a repeated symmetric Cournot duopoly in which a player's one-shot payoff function is given by

$$m(x_t, y_t) = \max\{109 - x_t - y_t, 0\} \cdot x_t - x_t, \tag{1}$$

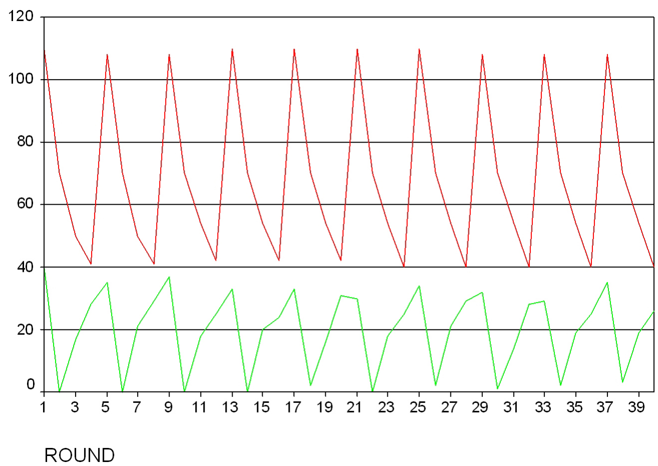
where $x_t \in \mathbb{R}_+$ (resp. $y_t \in \mathbb{R}_+$) denotes the action of the player (resp. opponent) in period t. Assume further that the opponent plays a myopic best response to the previous period's quantity of the player, that is

$$y_t = \max\left\{\frac{108 - x_{t-1}}{2}, 0\right\}. \tag{2}$$

Myopic best response can be viewed as a very simple adaptive heuristics. What is the player's optimal strategy against such an opponent? Is there a possibility to strategically manipulate the opponent such that he plays favorable to the player? This may require that the player forgoes some short-run profit in order to gain more in the long run.

We can view this setting as a dynamic programming problem for which the player's one period objective function is given by function (1) into which we substitute function (2). The problem is a bit non-standard in the sense that the objective function is not everywhere concave and differentiable, conditions usually required for dynamic programming (see Stokey, Lucas and Prescott, 1989, Bertsekas, 2005). Nevertheless, it is quite natural to conjecture that the optimal strategy of the player may involve to play a (current) best response in the last period and Stackelberg leadership in the previous periods. However, in an experiment in which human subjects played this game against a computer programmed to myopic best response (see Duersch, Kolb, Oechssler and Schipper, 2010), we discovered to our surprise one participant who played the 4-cycle of quantities depicted by the upper time series in Figure 1 and obtained a much higher average profit than the Stackelberg leader profit.[1] This experimental discovery triggered the current analysis. Can such a cycle be optimal?

Figure 1: Cycle played by a participant



In this article we will show that if the two-player game satisfies a version of strategic substitutes or strategic complements, namely decreasing or increasing differences, then the optimal control strategy is monotone in the initial action of the opponent, the discount rate, and in time periods. Examples of this class of games include some Cournot duopolies (Amir, 1996b), Bertrand duopolies (Vives, 1999), common pool resource games, public goods games, rent seeking games, Diamond's search, arms race (Milgrom and Roberts, 1990) etc. The key for the results is to apply methods from lattice programming (Topkis, 1978, 1998) to dynamic programming (see Topkis, 1978, Puterman, 1994, Amir, 1996a). It turns out that our problem is similar to a Ramsey-type capital accumulation problem solved in Amir (1996a). That paper was motivated by a very different application in

[1]The game was repeated over 40 rounds. The participant played the cycle of quantities (108, 70, 54, 42). This cycle yields an average payoff of 1520 which is well above Stackelberg leader payoff of 1458. In this game, the Stackelberg leader's quantity is 54, the follower's quantity is 27 (payoff 728), the Cournot Nash equilibrium quantity 36 (payoff 1296). The computer is programmed to myopic best response with some noise. The x-axis in Figure 1 indicates the rounds of play, the y-axis the quantities. The lower time series depicts the computer's sequence of actions. The upper time series shows the participant's quantities. See Duersch, Kolb, Oechssler and Schipper (2010) for details of the game and the experiment.

macroeconomics and no connection was made to our game theoretic problem. We apply his results to manipulation of myopic best-response learning and slightly extend it whenever necessary. Note that above example of the Cournot duopoly does not satisfy decreasing or increasing differences everywhere, which is caused by insisting on a non-negative price (see Section 3). That is, the results in Section 2 cannot be directly applied to our Cournot duopoly. Yet, we show in Section 3 how to use the results "piecewise" to conclude that a cycle of the four quantities $(108, 68, 54, 41)$ is the optimal control strategy, which is very close to the cycle $(108, 70, 54, 42)$ actually played by the participant in the experiment discussed above.[2]

Our approach in this paper bears some resemblance with the literature on infinitely repeated games with long-run and short-run players (sometimes referred to also as long-lived and short-lived players) (see Fudenberg, Kreps, and Maskin, 1990, Fudenberg and Levine, 1989, 1994). In this literature a long-run optimizer faces a sequence of static (or *current period's*) best response players who play only once. This is different from our model, in which the short-run player plays a best response to the *previous period's* action of the opponent. Our study can be seen as replacing the short-run player by a previous period's best response player. In a sense we "merge" the literature on repeated games with the literature on adaptive learning. As Fudenberg and Levine (1998, Chapter 8.11) point out, strategic teaching has been studied in repeated games with rational players but it is less prominent in learning theory. Camerer, Ho, and Chong (2002, 2006) study adaptive experience-weighted attraction learning of players in repeated games but allow for sophisticated players who respond optimally to their forecasts of all others' behavior. Their focus is on estimating such learning models with experimental data. There are only a few theoretical papers on learning in games in which players follow different learning theories (Banerjee and Weibull, 1995, Droste, Hommes, and Tuinstra, 2002, Hehenkamp and Kaarbøe, 2008, Juang, 2002, Schipper, 2009, 2017, Duersch, Oechssler, and Schipper, 2012, 2014). They focus on the evolutionary selection or relative success of different boundedly rational learning rules.[3] For instance, Duersch, Oechssler, and Schipper (2012) characterize the class of symmetric two-player games (that includes the Cournot duopoly mentioned in our paper) in which imitate-if-better cannot be beaten by any other decision rule no matter how sophisticated. Similarly, Duersch, Oechssler, and Schipper (2014) show that in symmetric two-player game tit-for-tat cannot be beaten by any other decision rule if and only if the game is an exact potential game. Another paper related to our work is Ellison (1997), who analyzes a large population which besides of players following a version of fictitious play also contains a single rational player. He shows that if players are randomly matched to play a 2x2 coordination game, the rational player may shift the play from a risk and Pareto dominated equilibrium to a risk

[2]In fact, the average payoff of the optimal cycle is 1522, only a minor improvement over the average payoff (1520) of the cycle played by the participant.

[3]As a reviewer pointed out, this literature is related to the literature on indirect evolution (e.g., Güth and Peleg, 2001, Heifetz, Shannon, and Spiegel, 2007). Yet, instead of the evolution of utility function, the evolution of learning heuristics is featured.

and Pareto dominant equilibrium but not vice versa. He also presents examples of some other 2x2 games and some 3x3 games, and shows that there can be cycles in which the rational player can achieve larger payoffs than in equilibrium. Our results go beyond 2x2 and 3x3 games and focus on a dynamic optimizer against a myopic best response player. In Section 4, we provide a further discussion of selected related literature.

The next section presents the model and monotonicity results. In Section 3 we discuss the cyclic Cournot example. We conclude with a discussion in Section 4. For better readability, all proofs are relegated to the appendix.

2 Model

2.1 The Dynamic Programming Problem

There are two players, a *manipulator* and a *puppet*. Let X, Y be two nonempty compact subsets of $\mathbb{R}$. We denote by $x \in X_y$ (resp. $y \in Y$) the manipulator's (resp. puppet's) action, where X_y is a continuous nonempty compact-valued correspondence from Y to 2^X. That is, we allow that the manipulator's set of actions may depend upon the puppet's action.[4]

Let $m : X \times Y \longrightarrow \mathbb{R}$ (resp. $p : Y \times X \longrightarrow \mathbb{R}$) be the manipulator's (resp. puppet's) one-period payoff function. We write $m(x_t, y_t)$ for the payoff obtained by the manipulator in period t if he plays x_t and the puppet plays y_t (analogous for the puppet). We assume that each player's payoff function is bounded. Further, we assume that m is upper semi-continuous on $X \times Y$ and p is continuous on $Y \times X$ and strictly quasiconcave in $y \in Y$. These assumptions are motivated in Lemma 1 below. Some of the assumptions may be stronger than necessary. For instance, in Section 4, we discuss how to weaken strict quasiconcavity of p to quasiconcavity. Note that we do not impose any concavity assumption on m.

Let $B : X \longrightarrow 2^Y$ be the puppet's best response correspondence. Moreover, let the puppet's best response function $b : X \longrightarrow Y$ be a selection of the best response correspondence, i.e., $b(x) \in B(x)$ for any $x \in X$.

Time is discrete and indexed by $t = 0, ..., T$. T may be infinity. We assume that the puppet is a myopic best response player with a given best response function. That is, given the manipulator's action x_{t-1} in period $t - 1$, the puppet's action at period t is

$$y_t = b(x_{t-1})$$

for $t = 1, ...$ and given $y_0 \in Y$.

Let $b(X)$ be the range of the puppet's best response function. We assume that $y_0 \in b(X)$, i.e., the puppet's initial action is a best response to some action of the

[4] In Section 4 we explain why we do not consider here multi-dimensional strategy sets.

manipulator. We believe that this assumption is not restrictive since a best response player should play by definition a best response to some action of the opponent.[5]

We allow the manipulator's set of actions to depend on the puppet's action for the sake of generality. The manipulator anticipates how his action today affects his set of feasible actions tomorrow via the puppet's best response tomorrow.[6] Nothing changes in our analysis if we were to assume instead that X_y is constant in y, but assumptions on X_y do play a role in the statements of the results.

For the existence of an optimal strategy, the manipulator's objective function should satisfy some continuity properties. While m above is assumed to be upper semicontinuous on $X \times Y$, this property does not necessarily extend to the modified one-period objective function $\hat{m}(\cdot,\cdot) := m(\cdot, b(\cdot))$ defined on $X \times X$. This is the reason for imposing stronger assumptions on the puppet's objective function p.[7] The following lemma is useful for the study of the optimization problem of the manipulator when the puppet is a myopic best response player.

Lemma 1 *If X_y is a continuous, nonempty, and compact-valued correspondence from Y to 2^X, m is upper semicontinuous on $X \times Y$, and p is continuous and strictly quasi-concave in y on Y given $x \in X$, then $\hat{m}(\cdot,\cdot) := m(\cdot, b(\cdot))$ is upper semicontinuous on $X \times X$ and $X_x := X_{b(x)}$ is a upper hemicontinuous, nonempty, and compact-valued correspondence from X to 2^X.*

The proof is contained in the appendix.

We can now consider the following Ramsey-type dynamic optimization problem

$$\sup \sum_{t=0}^{T-1} \delta^t \hat{m}(x_t, x_{t-1}) \tag{3}$$

s.t. $x_{-1} \in b^{-1}(y_0)$ given y_0, and $x_t \in X_{x_{t-1}}$ for $t = 0, 1, ..., T-1$, and $0 < \delta < 1$.

By standard arguments of dynamic programming (see Stokey, Lucas and Prescott, 1989, Bertsekas, 2005), the value function or Bellman equation satisfies

$$M_n(x) = \sup_{z \in X_x} \{\hat{m}(z, x) + \delta M_{n-1}(z)\} \tag{4}$$

for $n = 1, 2, ...$ with $M_0 \equiv 0$, and

$$M_\infty(x) = \sup_{z \in X_x} \{\hat{m}(z, x) + \delta M_\infty(z)\}. \tag{5}$$

[5]Note that throughout the analysis we do not allow the manipulator to choose suitably the initial action of the puppet.

[6]As a reviewer rightfully points out this would be problematic if the manipulator does not know the learning heuristic used by the puppet.

[7]As a reviewer pointed out, we could have stated the model just in terms of assumptions on m and a continuous best response function b. This might be even more realistic as the manipulator may observe the opponent's best responses but not necessarily the opponent's payoff function.

Note that the index in the equations corresponds to the time horizon of the optimization problem. $M_n(x)$ denotes the manipulator's objective function of the n-period dynamic optimization problem. That is, the index n runs backwards in time.

Lemma 2 *If X_y is a continuous, nonempty, and compact-valued correspondence from Y to 2^X, m is upper semicontinuous on $X \times Y$, and p is continuous and strictly quasi-concave in y on Y given $x \in X$, then for $n = 1, 2, ...,$ the value functions M_n and M_∞ are upper semicontinuous on X*

The proof is contained in the appendix.

In light of Lemma 2, optimal control strategies exist. We can replace the sup in equation (4) and (5) by the max. Let $S_n(x)$ be the arg max w.r.t. equation (4) (resp. (5)) if n is finite (resp. infinite). $S_n(x)$ is the set of all optimal decisions in the first period when the problem's horizon consists of n periods. Let s_n be a selection of S_n, and $\bar{s}_n$ and $\underline{s}_n$ be the maximum and minimum selection of S_n. If T is finite, we restrict attention to Markovian control strategies defined as sequence of transition functions $(d_0, d_1, ..., d_{T-1})$ with $d_t : X \longrightarrow X$ and $d_t(x) \in X_x$. When T is infinity, then we restrict our analysis to stationary Markovian control strategies $(d, d, ...)$ with $d : X \longrightarrow X$ and $d(x) \in X_x$. Such optimal control strategies exist but there may exist other optimal control strategies as well.

One may reasonably conjecture that actions of the manipulator approach the action of the leader in a Stackelberg outcome. This is not necessarily the case. This is easiest seen in the finite horizon. Suppose that the manipulator had played the Stackelberg leader quantity in the second last period. Then in the last period the puppet plays the Stackelberg follower quantity. Consequently, the optimal action of the manipulator in the last period is to play a best response to the Stackelberg follower quantity rather than to play the Stackelberg leader quantity. An example is discussed in the next section.[8] Nevertheless, there is a close connection to Stackelberg outcomes in terms of payoffs. The manipulator can guarantee herself the Stackelberg leader payoff (except for the initial period). The intuition is simply that since the puppet plays a best response to the manipulator's previous action, it would adjust to the Stackelberg follower quantity (with respect to his selection b from the best response correspondence) if the manipulator plays a Stackelberg leader quantity. Thus, the manipulator can now guarantee herself the Stackelberg leader payoff by simply playing the Stackelberg leader action.

To make this precise, let $X^S = \arg\max_{x \in X_x} m(x, b(x))$. This is the set of Stackelberg leader actions when the puppet uses best response selection b. Clearly, by Lemma 1, $X^S \neq \emptyset$. Assume any initial action of the puppet $y_0 \in Y$ such that $X^S \cap X_{y_0}$. This assumption avoids the problem that none of the Stackelberg leader actions are feasible when the puppet plays initial action y_0. Since there could be more than one Stackelberg

[8]In the first four periods, the cyclic example of Section 3 coincides with the smooth problem that we discuss in Section 3. Proposition 1 applies to this smooth problem. The manipulator's quantity in the last period is 41, which is the best response to the puppet's Stackelberg follower quantity.

leader action, let $x^s \in \arg\max_{x \in X^S \cap X_{y_o}} m(x, y_0)$. That is, x^s is a Stackelberg leader action that is feasible and payoff maximizing given the puppet's initial action y_0. Again, such action exists in light of Lemma 1. Now, given δ and the initial action of the puppet, y_0, define the n-period discounted Stackelberg leader payoff by

$$L_n(y_0) := m(x^s, y_0) + \sum_{t=1}^{n-1} \delta^t \hat{m}(x^s, x^s).$$

We can now state the observation on the payoff bound as follows:

Remark 1 *For any initial action of the puppet $y_0 \in Y$ such that $X^S \cap X_{y_0} \neq \emptyset$ and any time horizon $n > 1$, there exists $\bar{\delta} \in [0, 1)$ such that for all $\delta \in (\bar{\delta}, 1)$ and $x \in b^{-1}(y_0)$,*

$$M_n(x) \geq L_n(y_0) \text{ and } M_\infty(x) \geq L_n(y_0).$$

The proof follows now simply from the fact that the manipulator could resort to play the Stackelberg leader action x^s in every period, which would guarantee her a payoff of $L_n(y_0)$. The example discussed in the introduction and Section 3 shows that this bound is not necessarily tight. That is, there are games for which the manipulator can achieve a strictly higher payoff than the discounted Stackelberg leader payoff (see Footnote 1).

2.2 Monotonicity of Objective Functions

Before we can study properties of the solution for our dynamic optimization problem, we need to state some definitions and preliminary results. The first definition concerns a common notion of *strategic complements* (resp. *strategic substitutes*). A function $f : X \times Y \longrightarrow \mathbb{R}$ has *increasing (resp. decreasing) differences* in (x, y) on $X \times Y$ if for $x'' > x'$, $x'', x' \in X_{y''} \cap X_{y'}$ and for all $y'', y' \in Y \cap Y$ with $y'' > y'$,

$$f(x'', y'') - f(x', y'') \geq (\leq) f(x'', y') - f(x', y').$$

This function has *strictly increasing (resp. strictly decreasing) differences* if the inequality holds strictly. The function f is a *valuation* if it has both increasing and decreasing differences. The function f has *strongly increasing (resp. strongly decreasing) differences* in (x, y) on $X_y \times Y$ if $X, Y \subseteq \mathbb{R}_+$, X_y is a continuous, convex- and compact-valued correspondence from Y to 2^X, f is continuously differentiable, and for all $y'', y' \in Y$ with $y'' > y'$,

$$\frac{\partial f(x, y'')}{\partial x} > (<) \frac{\partial f(x, y')}{\partial x}.$$

A payoff function has *positive (resp. negative) externalities* if it is increasing (resp. decreasing) in the opponent's action.

The following examples illustrate the relatively broad applicability of our analysis. More general versions could be considered but the point is that many any "textbook"-like two-player games from a wide array of applications fall into the class of games we consider here.

Example 1 (Cournot Duopoly with Linear Demand) *Consider a (quasi) Cournot duopoly given by the symmetric payoff function $m(x, y) = (b - x - y)x - c(x)$ with $b > 0$ (and analogously for $p(y, x)$). The payoff function m has strictly decreasing differences and negative externalities. Moreover, p is strictly concave if costs are convex.*

Example 2 (Bertrand Duopoly with Product Differentiation) *Consider a differentiated duopoly with constant marginal costs, in which the manipulator's payoff function is given by $m(x, y) = (x - c)(a + by - \frac{1}{2}x)$, for $a > 0$, $b \in [0, 1/2)$ (and symmetrically for the puppet). This game has strictly increasing differences and positive externalities. Moreover, payoffs are also strictly concave.*

Example 3 (Public Goods) *Consider the class of symmetric public good games defined by $m(x, y) = g(x, y) - c(x)$ where $g(x, y)$ is some symmetric benefit function increasing in both arguments and $c(x)$ is an increasing cost function (and symmetrically for p). Various assumptions on g have been studied in the literature but often some complementarity of contributions is assumed. If g has increasing differences then so has m and p. If g is concave in y and c is convex, then p is convex.*

Example 4 (Common Pool Resources) *Consider a common pool resource game with two appropriators (Walker, Gardner, and Ostrom, 1990). Each appropriator has an endowment $e > 0$ that can be invested in an outside activity with marginal payoff $c > 0$ or into the common pool resource. Let $x \in X \subseteq [0, e]$ denote the opponent's investment into the common pool resource (likewise y denotes the imitator's investment). The return from investment into the common pool resource is $\frac{x}{x+y}(a(x + y) - b(x + y)^2)$, with $a, b > 0$. So the manipulator's payoff function is given by $m(x, y) = c(e - x) + \frac{x}{x+y}(a(x+y) - b(x+y)^2)$ if $x, y > 0$ and ce otherwise (and symmetrically for the puppet). This game has negative externalities, decreasing differences, and is strictly concave in the player's action.*

Example 5 (Minimum Effort Coordination) *Consider the class of minimum effort games given by the symmetric payoff function $m(x, y) = \min\{x, y\} - c(x)$ for some convex cost function $c(\cdot)$ (see Bryant, 1983, and Van Huyck, Battalio, and Beil, 1990). This game has positive externalities, increasing differences, and is concave.*

Example 6 (Synergistic Relationship) *Consider a synergistic relationship among two individuals. If both devote more effort to the relationship, then they are both better off, but for any given effort of the opponent, the return of the player's effort first increases and then decreases (see Osborne, 2004, p.39). The manipulator's payoff function is given* *by $m(x, y) = x(c + y - x)$ with $c > 0$ and $x, y \in X \subset \mathbb{R}_+$ with X compact (and symmetrically for the puppet). This game has positive externalities, increasing differences, and is strictly concave.*

Example 7 (Diamond's Search) *Consider two players who exert effort searching for a trading partner. Any trader's probability of finding another particular trader is proportional to his own effort and the effort by the other. The symmetric payoff function is given by $m(x, y) = \alpha xy - c(x)$ for $\alpha > 0$ and c increasing and convex (see Milgrom and Roberts, 1990, p. 1270). The game has positive externalities, increasing differences, and is concave.*

A set of actions $X_y \subseteq \mathbb{R}$ is *expanding (resp. contracting)* if $y'' \geq y'$ in Y implies that $X_{y''} \supseteq (\subseteq) X_{y'}$. A *correspondence* $F : X \longrightarrow 2^Y$ is *increasing (resp. decreasing)* if $x'' \geq x'$ in X, $y'' \in F(x'')$, $y' \in F(x')$ implies that $\max\{y'', y'\} \in F(x'')$ (resp. $\min\{y'', y'\} \in F(x')$).

The following lemma shows how above conditions on the game's payoff functions m and p translate into properties of the manipulator's objective function $\hat{m}$. These properties will allow us later on to show properties of optimal control strategies.

Lemma 3 (Properties of $\hat{m}$) *(i) Monotone Differences: Table 1 establishes relationships between increasing and decreasing differences of m, p, and $\hat{m}$. E.g., if both m and p have increasing differences, then so has $\hat{m}$ (first line of Table 1).*

(ii) Monotonicity in the Second Argument: Table 2 establishes relationships between positive and negative externalities of m, increasing or decreasing differences of p, and monotonicity of $\hat{m}(x_{t+1}, x_t)$ in x_t. E.g., if m has positive externalities and p has increasing differences, then $\hat{m}(x_{t+1}, x_t)$ is increasing in x_t for every x_{t+1} (first line of Table 2).

The proof is contained in the appendix. It makes use of results by Topkis (1998).

According to Lemma 3 (i) whenever m and p have the same kind of monotone differences, then $\hat{m}$ has increasing differences. Moreover, when m and p have different kinds of monotone differences, then $\hat{m}$ has decreasing differences. These facts are not too surprising. Monotone differences of the puppet's objective function translate into monotone best responses. If the puppet's best responses are increasing - as in the case of increasing differences of the puppet's payoff function - then it preserves increasing differences of the manipulator's payoff function (understood now as a function of the manipulator's actions today and yesterday). Yet, with decreasing puppet's best responses, it essentially reorders the second argument of the manipulator's payoff function by the dual order and the manipulator's payoff function with increasing differences in her and the opponent's actions has now decreasing differences in her actions today and yesterday.

The significance of Lemma 3 is that it allows us now to apply some results proved for Ramsey-type capital accumulation problems by Amir (1996a) (see Puterman, 1994, for

Table 1:

If m has				**and** p has				**then** $\hat{m}$ has			
strongly	strictly	incr.	decr	strongly	strictly	incr.	decr.	strongly	strictly	incr.	decr.
		differences				differences				differences	
		√				√				√	
			√				√			√	
		√					√				√
			√			√					√
	√	√		√	√	√			√	√	
	√		√	√	√		√		√	√	
	√	√		√	√		√		√		√
	√		√	√	√	√			√		√
√	√	√		√	√	√		√	√	√	
√	√		√	√	√		√	√	√	√	
√	√	√		√	√		√	√	√		√
√	√		√	√	√	√		√	√		√

Table 2:

If m has		**and** p has		**then** $\hat{m}(x_{t+1}, x_t)$ is	
positive	negative	increasing	decreasing	increasing	decreasing
externalities		differences		in x_t	
√		√		√	
√			√		√
	√	√			√
	√		√	√	

related results) to derive some properties of n-period value functions. Lemma 4 states that the n-period value functions are monotone in the previous period's action of the manipulator.

Lemma 4 *Table 3 establishes the monotonicity of the n-period value functions M_n in the previous period's action of the manipulator. E.g., if m has positive externalities, p has increasing differences, and X_y is expanding, then M_n is increasing in x (first line of Table 3).*

The proof follows from above lemmata and the proof of Theorem 1(i) in Amir (1996a).

Table 3:

If m has		**and** p has		**and** X_y is		**then** M_n is on X	
positive	negative	increasing	decreasing	expanding	contracting	increasing	decreasing
externalities		differences					
√		√		√		√	
	√		√	√		√	
√			√		√		√
	√	√			√		√

According to Lemma 4 the monotonicity of the n-period value function does not depend on increasing or decreasing differences of the manipulator's payoff function. This is because it pertains to the monotonicity with respect to *previous* period's action by the manipulator. This also suggests that it should depend crucially on whether m displays positive or negative externalities in the puppet's action.

2.3 Monotone Optimal Control Strategies

Proposition 1 (i) states that the n-period optimal control strategies are monotone in the previous period's action of the manipulator. The monotonicity crucially depends on both the increasing or decreasing differences of m and p. Whenever, both m and p have increasing differences or both have decreasing differences, then largest and smallest selections of n-period optimal control strategies are increasing in the previous period's action of the manipulator. Otherwise, if monotone differences of m and p differ, these selections are decreasing in the previous period's action of the manipulator. In analogy to monotone best responses for games with monotone differences, this can be viewed as a results on monotone "dynamic" best responses (i.e., monotone best responses to a dynamic optimization problem).

Proposition 1 (ii) claims that the $(n+1)$-horizon optimal control strategy is larger than the n-horizon optimal control strategy. That is, optimal control strategies are monotone over time. This means, that for any initial action of the manipulator the first-period action of the manipulator in the $n+1$-period horizon optimization problem is larger than the manipulator's first-period action in the n-period horizon problem. This can be viewed as a dynamic version of Topkis style monotone comparative statics result where the comparative statics is with respect to the time horizon of the dynamic optimization problem.

Finally, Proposition 1 (iii) states sufficient conditions for maximal and minimal selections of the optimal control strategies being monotone increasing in the discount rate.

This is a dynamic version of Topkis style monotone comparative statics results where the comparative statics is with respect to the discount factor.

Proposition 1 *(i) Table 4 establishes that the n-period optimal control strategies are monotone in the previous period's action of the manipulator, for $n = 1, 2, \ldots$ E.g., if both m and p have increasing differences and X_y is increasing, then both $\bar{s}_n$ and $\underline{s}_n$ is increasing on X (first line of Table 4).*

Table 4:

If m has			**and** p has			**and** X_y is		**then** ... is	
strictly	incr.	decr.	strongly	incr.	decr.	increasing	decreasing	incr.	decr.
differences			differences					on X	
	√			√		√		$\bar{s}_n$, $\underline{s}_n$	
		√			√	√		$\bar{s}_n$, $\underline{s}_n$	
	√				√		√		$\bar{s}_n$, $\underline{s}_n$
		√		√			√		$\bar{s}_n$, $\underline{s}_n$
√	√		√	√		√		s_n	
√		√	√		√	√		s_n	
√	√		√		√		√		s_n
√		√	√	√			√		s_n

(ii) Table 5 establishes the relationship between $(n+1)$-horizon optimal control strategy and the n-horizon optimal control strategy, for $n = 1, 2, \ldots$. E.g., if m has positive externalities, both m and p have increasing differences, and X_y is expanding, then for $n = 1, 2, \ldots$, $\bar{s}_{n+1} \geq \bar{s}_n$ and $\underline{s}_{n+1} \geq \underline{s}_n$ (first line of Table 5).

Table 5:

If m has		**and** both m and p have		**and** X_y is		**then**
positive	negative	increasing	decreasing	expanding	contracting	for $n = 1, 2, \ldots$
externalities		differences				
√		√		√		$\bar{s}_{n+1} \geq \bar{s}_n$, $\underline{s}_{n+1} \geq \underline{s}_n$
	√		√	√		$\bar{s}_{n+1} \geq \bar{s}_n$, $\underline{s}_{n+1} \geq \underline{s}_n$
√			√		√	$\bar{s}_{n+1} \leq \bar{s}_n$, $\underline{s}_{n+1} \leq \underline{s}_n$
	√	√			√	$\bar{s}_{n+1} \leq \bar{s}_n$, $\underline{s}_{n+1} \leq \underline{s}_n$

(iii) *Suppose that [m has positive externalities and both m and p have increasing differences] or [m has negative externalities and both m and p have decreasing differences] and* X_y *is expanding. If* $\delta'' \geq \delta'$, $\delta'', \delta' \in (0,1)$, *then* $\bar{s}_n(\cdot, \delta'') \geq \bar{s}_n(\cdot, \delta')$ *and* $\underline{s}_n(\cdot, \delta'') \geq \underline{s}_n(\cdot, \delta')$.

This proposition is essentially an application of Topkis's (1978, 1998) results on the monotone comparative statics of supermodular functions on lattices. The proof of Proposition 1 (i) follows from above lemmata and the proof of Theorem 1 (iii) in Amir (1996a). The proof of the first two lines in the table of Proposition 1 (ii) follow from above lemmata and Amir (1996a, Theorem 2 (i)). The last two lines extend Theorem 2 (i) in Amir (1996a) and the proof is contained in the appendix. Such extension becomes possible here because we focus only on single-dimensional variables whereas Amir (1996a) allows the set of variables to be a lattice. Given previous lemmata, the proof of Proposition 1 (iii) is essentially analogous to the proof of Theorem 2 (ii) in Amir (1996a). Nevertheless we decided to state it in the appendix.[9]

One may be tempted to conjecture analogous results to Proposition 1 (ii) for cases in which the monotone differences of m and p differ. In the appendix we show an auxiliary result (Proposition 3) according to which if monotone differences of payoff functions differ, then $M_n(x)$ has *no monotone differences* in (n, x) unless it is a valuation. Hence, we can not hope to prove with the same methods a result similar to Proposition 1 (ii) if monotone differences of m and p differ. How do optimal control strategies look like in such cases? Below Example 1 suggests that if monotone differences of m and p differ, then the optimal control strategy may involve a cycle. Moreover the example shows that the manipulator may play a strictly dominated action of the one-shot game within the cycle. Thus, apparent "irrational" behavior may in fact be rational in a dynamic context even if just finite repetitions are considered.

Example 8 *Consider the following 2x2 game:*

	l	r
t	$0,1$	$0,3$
d	$6,6$	$20,4$

For any possible ordering of each player's action set, the game has monotone differences but the monotone differences differ among players. That is, if either [$l > r$ *and* $t > d$*] or [*$l < r$ *and* $t < d$*], then the row player's payoff function has increasing differences whereas the column player's payoff function has decreasing differences. Otherwise, if either [*$l > r$ *and* $t < d$*] or [*$l < r$ *and* $t > d$*], then the row player's payoff function has decreasing differences whereas the column player's payoff function has increasing differences.*

Let the manipulator's payoff function correspond to the row player's payoffs, and the puppet's payoff function to the column player's payoffs. If $T \geq 2$, T *an even integer (T*

[9] Amir (1996a, Theorem 2 (ii)) does not state explicitly that the one-period value function is increasing and X_y is expanding. Yet, this property is required in the proof.

may be finite), then it is easy to see that a cycle of $t, d, t, d, ...$ *is optimal. If the puppet's initial action is* l*, such a cycle yields a payoff stream of* $0, 20, 0, 20, ...$ *whereas repeated play of the unique Nash equilibrium action* $d, d, d, d, ...$ *yields* $6, 6, 6, 6, ...$

Note that t *is strictly dominated by* d *in the one-shot game. Thus, the example demonstrates that the manipulator may use a strictly dominanted action in an optimal control strategy if it induces the puppet to a response favorable to the manipulator.*[10]

One may also conjecture results analogous to Proposition 1 (iii) for the cases in which monotone differences of m and p differ or when externalities of m are reversed. We discuss this in the appendix and show with some auxiliary results (Proposition 4) that the results of Proposition 1 do not extend to such cases.

The next proposition strengthens the conclusions of Proposition 1 to strict monotonicity. This comes at the cost of assuming strongly increasing or decreasing differences (and thus the differentiability of the payoff functions). The result may be useful in applications where strict monotonicity is of interest.

Proposition 2 *Let* X *be a nonempty, convex compact subset of* $\mathbb{R}_+$*, and let* X_x *be a compact-valued, convex-valued, and continuous correspondence from* X *to* 2^X*. Moreover, let* s_n *be any interior optimal strategy for* $n = 1, 2, ...$*, i.e.* $s_n(x)$ *is in the interior of* X_x*.*

(i) Table 6 establishes that the n-period optimal control strategies are monotone in the previous period's action of the manipulator, for $n = 1, 2, ...$ *E.g., if both* m *and* p *have strongly increasing differences and* X_y *is increasing, then* s_n *is strictly increasing on* X *(first line of Table 6).*

Table 6:

If m has strongly		**and** p has strongly		**and** X_y is		**then** ... is strictly	
incr.	decr.	incr.	decr.	incr.	decr.	incr.	decr.
differences		differences				on X	
$\surd$		$\surd$		$\surd$		s_n	
	$\surd$		$\surd$	$\surd$		s_n	
$\surd$			$\surd$		$\surd$		s_n
	$\surd$	$\surd$			$\surd$		s_n

(ii) Table 7 establishes the relationship between $(n + 1)$*-horizon optimal control strategy and the n-horizon optimal control strategy, for* $n = 1, 2, ...$ *E.g., if* m *has*

Table 7:

If m has		**and** both m and p have strongly		**and** X_y is		**then**
positive	negative	increasing	decreasing	expanding	contracting	for $n = 1, 2, ...$
externalities		differences				
$\surd$		$\surd$		$\surd$		$s_{n+1} > s_n$
	$\surd$		$\surd$	$\surd$		$s_{n+1} > s_n$
$\surd$			$\surd$		$\surd$	$s_{n+1} < s_n$
	$\surd$	$\surd$			$\surd$	$s_{n+1} < s_n$

positive externalities, both m and p have strongly increasing differences, and X_y is expanding, then $n = 1, ...,$ $s_{n+1} > s_n$ (first line of Table 7).

(iii) Suppose that [m has positive externalities and both m and p have strongly increasing differences] or [m has negative externalities and both m and p have strongly decreasing differences] and X_y is expanding. If $\delta'' > \delta'$, $\delta'', \delta' \in (0, 1)$, then $s_n(\cdot, \delta'') > s_n(\cdot, \delta')$.

The proofs of the first two lines in Proposition 2 (i) follow from previous lemmata and Amir (1996a, Theorem 3(i)). The last two lines extend Amir (1996a, Theorem 3(i)), and the proof is contained in the appendix. Such an extension becomes possible here because we focus on one-dimensional action sets only. The proof of the first two lines in Proposition 2 (ii) follow from previous lemmata and Amir (1996a, Theorem 3 (ii)). The last two lines extend Amir (1996a, Theorem 3(ii)), and the proof is contained in the appendix. Again, such an extension becomes possible here because we focus on one-dimensional action sets only. The proof of Proposition 2 (iii) follows from previous lemmata, Proposition 1 (iii), and the proof of Amir (1996a, Theorem 3 (iii)).

3 The Cyclic Example

Consider the standard textbook Cournot duopoly discussed in the introduction. In this section we want to show that a cycle is optimal in this example. First note that the results from the previous section *do not* apply to the example. The Cournot duopoly does not satisfy decreasing differences everywhere, which is due to insisting on a non-negative

[10] This finding that an optimal control strategy involve strictly dominated actions is not restricted to games for which monotone differences differ among players.

price. To see this note that for instance

$$\begin{aligned} m(100,0) - m(50,0) &< m(100,100) - m(50,100) \\ 800 - 2900 &< -100 - 50 \end{aligned}$$

while

$$\begin{aligned} m(40,20) - m(30,20) &> m(40,30) - m(30,30) \\ 1920 - 1740 &> 1520 - 1440. \end{aligned}$$

Consider now a "smooth" version of the game, in which we do not insist on a non-negative price. The symmetric payoff function is given by

$$\tilde{m}(x,y) = (108 - x - y) \cdot x.$$

This game has strongly decreasing differences everywhere and negative externalities. Thus, Proposition 1 applies. Optimal control strategies are monotone decreasing over time periods, decreasing in the puppet's initial quantity, and increasing in the discount factor. The graph of the smooth payoff function $\tilde{m}$ is identical to the graph of the original payoff function for the range of actions $x \in [0, 109 - y]$. In this range the original game satisfies strictly decreasing differences. Similarly, for any n we can find the range of x_0 where the smooth n-period objective function coincides with the original n-period's objective function.

We want to prove that a cycle of four actions $(108, 68, 54, 41)$ is optimal. This cycle is very close to the cycle actually played by the participant in the experiment as discussed in the introduction. The idea of the proof is as follows: Since we consider a finite repetition of the game, we can use backwards induction. By our previous results, any optimal sequence of actions must be monotonically decreasing over time as long as x_0 is in the range where the n-period objective function coincides with the smooth n-period objective function.[11] We show that if the game is repeated for eight periods, then this assumption is violated in the fifth period. We show that in this game it means that there must be cycle if $n = 8$, and it turns out that the 4-cycle $(108, 68, 54, 41)$ is optimal. Using our monotonicity results, we extend the result to $n > 8$.

For $n = 1, 2, ..., 8$, we write down recursively the n-period objective functions $M_n(x_1, x_0)$ as a function of the past quantity x_0 and the quantity decided by the manipulator in period 1. The quantities of the manipulator in the following $n-1$ periods are the optimal

[11] Since we look at cycles (of finite length), we can neglect discounting in the calculations below.

quantities given by the $n-1$ period problems.[12]

$$
\begin{aligned}
M_1(x_1,x_0) &= \max\{109 - x_1 - b(x_0), 0\} \cdot x_1 - x_1 \\
M_2(x_1,x_0) &= \max\{109 - x_1 - b(x_0), 0\} \cdot x_1 - x_1 \\
&\quad + \max\{109 - s_1(x_1) - b(x_1)\} \cdot s_1(x_1) - s_1(x_1) \\
M_3(x_1,x_0) &= \max\{109 - x_1 - b(x_0), 0\} \cdot x_1 - x_1 \\
&\quad + \max\{109 - s_2(x_1) - b(x_1)\} \cdot s_2(x_1) - s_2(x_1) \\
&\quad + \max\{109 - s_1(x_2) - b(x_2)\} \cdot s_1(x_2) - s_1(x_2) \\
\vdots \quad &\vdots \quad \vdots
\end{aligned}
$$

and solve for the n-period optimal control action $s_n(x)$ under the assumption that x is in the range where the n-period objective function coincides with the smooth n-period objective function:[13]

$$
\begin{aligned}
s_1(x) &= \frac{1}{4} \cdot x + 27 \text{ if } x \in [0, 108] \\
s_2(x) &= \frac{4}{15} \cdot x + 36 \text{ if } x \in [41.59, 108] \\
s_3(x) &= \frac{15}{56} \cdot x + \frac{270}{7} \text{ if } x \in [53.560, 108] \\
s_4(x) &= \frac{56}{209} \cdot x + \frac{432}{11} \text{ if } x \in [56.264, 108] \\
s_5(x) &= \frac{209}{780} \cdot x + \frac{513}{13} \text{ if } x \in [56.959, 108] \\
s_6(x) &= \frac{780}{2911} \cdot x + \frac{1620}{41} \text{ if } x \in [57.142, 108] \\
s_7(x) &= \frac{2911}{10864} \cdot x + \frac{3834}{97} \text{ if } x \in [57.191, 108] \\
s_8(x) &= \frac{10864}{40545} \cdot x + \frac{672}{17} \text{ if } x \in [57.204, 108]
\end{aligned}
$$

E.g., $s_2(x)$ above is the manipulator's optimal first period quantity to the two-period problem for the original non-smooth problem if the initial quantity satisfies $x \in [41.59, 108]$ since under the latter condition the non-smooth problem coincides with the smooth problem.

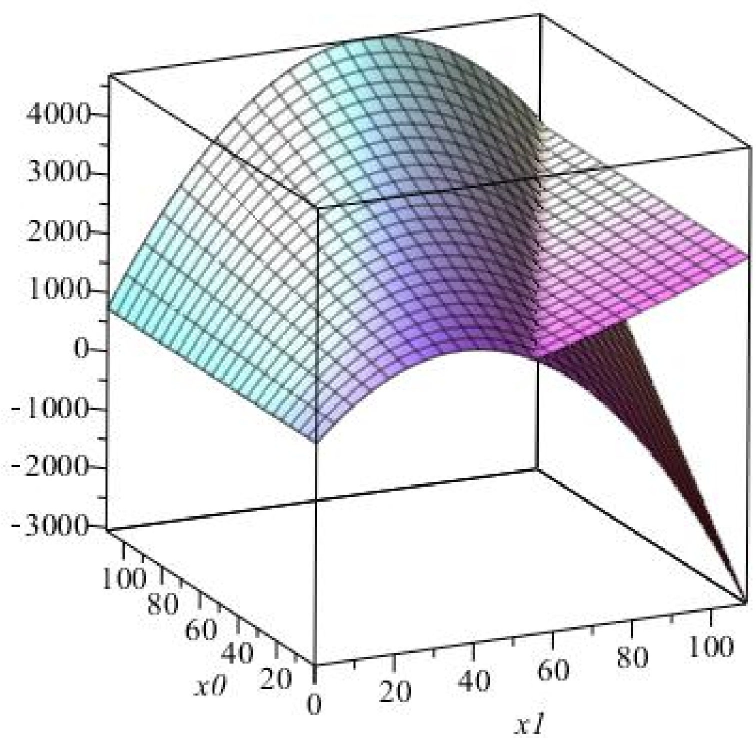
If x is outside the respective for range for which the n-period objective function coincides with the smooth n-period objective function, then there is a corner solution $s_n(x) = 108$ since the graph of the n-period objective function has the typical shape depicted in Figure 2. The figure depicts as an example the smooth (lower graph) and the

[12]To save space, we write out only the objective functions for $n = 1, 2, 3$.

[13]Interestingly, the denominator in the linear factor in s_n is identical the numerator of the linear factor in s_{n+1}.

original (upper graph) n-period objective functions for $n = 2$, $M_2(x_1, x_0)$. The vertical axis measures profits over two periods. Variable $x1$ refers to the decision variable of the manipulator in the first period of the 2-period problem. Variable $x0$ is the manipulator's initial quantity. For $n > 2$, the graphs of the objective functions are qualitatively similar. There is always a segment in which the smooth and original objective functions separate.

Figure 2: Smooth and Original Objective Functions for $n = 2$



If $x_1 = 108$, then $M_n(108, x_0)$ is constant for all $x_0 > 1$. That is, if $x_1 = 108$ then the n-periods payoff is constant in x_0. It does not matter to the manipulator what the puppet plays in period 1. In particular, the puppet could play a best response to x_0. Now, this creates an opportunity for a n-period cycle to emerge that starts with the manipulator playing $x_1 = 108$. Then by Proposition 1 the quantity of the manipulator would fall monotonically as time progresses as long as the $(n - k)$-period objective function coincides with the smooth $(n - k)$-period objective function for $k = 1, ..., n - 1$ since the game has decreasing differences and negative externalities. The cycle restarts once the manipulator's quantity falls outside the range for which the smooth n-period objective function coincides with the original n-period objective function. At this point, the manipulator plays again 108 and the puppet best responds to the previous quantity of the manipulator. The cycle length depends now on n for which the manipulator's quantity falls outside the range for which the smooth n-period objective function coincides with the original n-period objective function.

In the experiment mentioned in the introduction, the initial puppet's action was set to $y = 40$. That is, if we consider the $(n = 8)$-period problem, already in the 0-period's quantity $x_0 = 28$ (defined by $40 = b(x_0)$) would be outside the range for which the 8-period objective function coincides with the smooth 8-period objective function. Hence there must be at least an 8-cycle (or lower cycle-length) in the 8-period problem.

Suppose there is such a 8-cycle in the 8-period problem, then by above arguments $x_8 = 108$. Using the n-period optimal control strategies for $n = 1, 2, ..., 6, 7$ above, we can compute the optimal sequence of quantities of the manipulator when $x_8 = 108$:

t	1	2	3	4	5	6	7	8
n	8	7	6	5	4	3	2	1
$s_n(x_{t=8-n}) = x_t$	108	68.464	57.857	54.964	54	53.036	50.143	39.536

We note that $n = 4$ is the largest n for which the previous period's quantity x_{8-n} is outside the range for which the n-period objective function coincides with the n-period smooth objective function. I.e., $x_4 = 54.964 \notin [56.264, 108]$. Therefore we cannot use $s_4(x_4)$ to compute the optimal quantity in $t = 5$. Hence, x_5 above is not optimal. Thus, the proposed 8-cycle cannot be optimal. It follows that a smaller cycle must be optimal. Indeed, when we compute all smaller cycles using n-period optimal control strategies s_n with starting value 108, then we find that the 4-cycle is optimal.

Consider now the strategic control problems for this game with more than k periods for $k > n$ for k being a multiple of 4. Suppose that a 4-cycle is not optimal anymore for such a problem with a time period larger than 8. Then we must have that x_{k-4} in optimal path for the k-period problem is strictly larger than x_4 for the 8 period problem above. Otherwise, by previous arguments the 4-cycle would be optimal. This could only be true if x_{k-7} in the optimal path of k-period problem is strictly larger than x_1 for the 8-period problem, since by Proposition 1 (i) for $n = 1, 2, ...$ we have that s_n is monotone increasing in the previous period's quantity. However, already for the 8-period problem we have $x_1 = 108$, the largest undominated action that makes the puppet leave the market in the following period. Hence, x_{k-7} in the optimal path for the k period problem can not be larger than 108. This implies that x_{k-4} in optimal path for the k-period problem is not strictly larger than x_4 in the 8-period problem. We conclude that the 4-cycle is optimal.

What happens if there is a finite repetition of the game for which the number of periods can not be divided by 4? For all problems with less then 8 periods it is easy to verify that in the last 4 periods the 4-cycle is optimal. In any previous periods there is an optimal path monotone over periods since the range-assumption won't be violated. For problems with a finite number of periods larger than 8 that can not divided by 4, the 4-cycle is optimal for the last 4ℓ periods for $\ell = 1, 2, ...$ For any previous periods, there is an optimal path monotone over periods since the range-assumption is not violated.

The result of optimal cycles may be generalized to a larger class of Cournot games in which we insist on a non-smooth lower bound for the price although the optimal cycle length and quantities in the cycle may depend on the parameters of the game. Yet, quantities should decrease over the length of each cycle before jumping up again since the game has decreasing differences for substantial range of quantities as well as negative externalities. The importance of the non-smooth lower bound for prices is that for market quantities that would push the prices even lower, the game displays increasing differences instead of decreasing differences. That is, the monotone differences switch once market quantities are such that the price hits its lower bound. This is the point at which the graph of the original objective function differs from the graph of the smooth objective function in Figure 2. See the beginning of this section for an illustration.

Finally, we note that the example is not non-generic. That is, small perturbations of the payoff functions m and p do not change the result qualitatively.

4 Discussion

In this article we assumed that actions are one-dimensional although lattice programming allows usually to prove results even if strategies are multi-dimensional. The crucial assumption required is that payoffs are supermodular in actions. To see what may go wrong in our case, note that if we assume that both m and p are supermodular in actions, then $\hat{m}$ may not be supermodular even if every best response selection $b(x)$ is supermodular in x. E.g. the composition of $m(\cdot, -b(x))$ may not be supermodular in x on X.

We used the cardinal properties of decreasing and increasing differences to obtain our results. Our results can not necessarily be extended to the weaker ordinal notion of (dual) single crossing property. The reason is that the manipulator's objective function is a weighted sum of one-period payoff functions. It is well know that the sum of functions each satisfying the single-crossing property may not satisfy the single-crossing property (Milgrom and Shannon, 1994).

In Lemma 1 we assume that p is strict quasi-concave in y. This is probably too strong. We require that m is upper semicontinuous and b continuous, since if b is just upper semicontinuous the composition $\hat{m}$ may not be an upper semicontinuous function. E.g., if b is an upper semicontinuous function then $-b$ is a lower semicontinuous function. Hence $m(\cdot, -b(\cdot))$ may not be a upper semicontinuous function. It would suffice to obtain a continuous selection b from B. By Michael's Selection Theorem we could require that B is a convex-valued lower hemicontinuous correspondence. But the Theorem of the Maximum just yields a upper hemicontinuous correspondence. As a remedy, we could try to find an approximation along arguments similar to the one used in generalizing Brouwer's fixed point theorem to Kakutani's fixed point theorem. While it may not be possible to find a continuous selection of an upper hemicontinuous correspondence, a convex-valued upper hemicontinuous correspondence can be approximated by a closed and convex-valued lower hemicontinuous correspondence. Note that convex-valuedness of B requires quasi-concavity of p anyway.

In our model we require that the initial action of the puppet is a best response to some action of the manipulator. This may be quite restrictive when period 0 is viewed as the first period. After all a motivation for learning theories is to study whether boundedly rational learning could converge to a rational action without assuming that players start already with it. Yet, we believe that this assumption is not restrictive because myopic best response players are programmed to best replies. So no matter what they play, it should be a best response to some of the opponent's action.

At the first glance, the optimal cycle in the Cournot duopoly with a non-negative price may look surprising. Yet, we also found optimal cycles in games where one player's payoff function has increasing differences while the other player's payoff function has decreasing differences (Example 1). Moreover, it is easy to see that the optimal control strategy against a myopic best response player in a matching pennies game involves a two-cycle. Similarly, a three-cycle is optimal in the Rock-Paper-Scissors game. Note however that the optimal cycle in the Cournot game or Example 1 is more subtle since it involves the manipulator's play of strict dominanted actions of the one-shot game while in those zero-sum games the manipulator always plays a best response and hence he does not need to sacrifice short term for long term gain.[14] I am thankful to a reviewer pointing out that optimal cycles in control problems also appear in competitive economies with a representative agent (Benhabib and Nishimura, 1985, Boldrin and Montrucchio, 1986).

Any optimal cycles are due to the "mechanistic" nature of myopic best response. It seems quite unrealistic that a player even if he is adaptive should not recognize cycles after some time. Aoyagi (1996) studies repeated two-player games with adaptive players who are able to recognize patterns such as cycles in the path of play. Indeed, it may be worthwhile to extend our analysis and allow the best response player to recognize cycles.

One reviewer raised the interesting question whether an example of Cournot duopoly could be designed in which there is a chaotic optimal orbit (which would also preclude the benefit of cycle recognition). It is known in the literature that chaotic orbits can appear with best response type dynamics in versions of Cournot duopoly (Rand, 1978). Moreover, as another reviewer pointed out, chaos has been observed in optimal capital accumulation problems (Boldrin and Montrucchio, 1986). However, we conjecture that it is impossible to create an example of Cournot duopoly with a myopic best response player against whom the dynamically optimal player has an optimal chaotic strategy. The reason is that in the range of output combinations in which the game has decreasing differences, quantities of the optimizer are monotone decreasing over time (Proposition 1). So one would need to create an example without decreasing differences in some "significant" portion of the range. Such a game is hard to interpret as Cournot competition that

[14] We like to remark that not in all zero-sum games the optimal control strategy of the manipulator involves a cycle. This is the case for some classes of zero-sum games studied in Duersch, Oechssler and Schipper (2012, 2014).

almost epitomizes decreasing differences. To sum up, while it possible to create complex dynamics in Cournot duopoly, we doubt that there could be an example that still satisfies the main features of our cyclic example, namely one myopic best response player and one dynamic optimizer in Cournot duopoly.

Recently, Schipper (2017) studied strategic teaching of uncoupled learning heuristics leading to Nash equilibrium. A learning heuristic is uncoupled if it does not take as input opponents' payoffs. Myopic best response is one example of an uncoupled learning heuristic. It also converges to Nash equilibrium in an interesting class of games that contains our example of the Cournot duopoly. Previous research has shown that there exist uncouple learning heuristics that converge to Nash equilibrium in all finite games. Schipper (2017) shows that there does not exist an uncouple learning learning heuristic that if played by all players converges to Nash equilibrium in all finite games but that cannot be manipulated away from Nash equilibrium. Some player may teach opponents in order to increase his long run payoff. This result applies even when restricted to the class of games with increasing differences or decreasing differences as considered in our paper. His results are very general as they apply to large classes of games and all uncoupled learning heuristics leading to Nash equilibrium. The flip-side is that he is unable characterize the optimal strategic teaching strategy. In the current paper we focus on one uncoupled learning learning heuristic and particular interesting classes of games and characterize the optimal strategic teaching strategy. Schipper (2017) also characterizes bounds on long run payoff that a manipulator can achieve against an uncoupled learning heuristic leading to Nash equilibrium. He shows that the manipulator can achieve at least the worst Stackelberg leader payoff. This generalizes Remark 1 to all uncoupled learning heuristics converging to Nash equilibrium in all games or subclasses of games I consider in this paper.

A related recent paper is Kordonis, Charalampidis, and Papavassilopoulos (2017). They study games with uncertainty over payoffs so that players can try to manipulate other players by pretending to have payoffs different from their actual payoffs. With only one pretender, she can achieve payoffs of the Stackelberg leader, echoing the literature on reputation formation in repeated games (e.g., Fudenberg and Levine, 1989) and strategic teaching of learners in Schipper (2017). Several classes of games are considered and the results are applied to electricity markets.

We view our analysis as a first step towards studying strategic control of a particular adaptive learning heuristic. We envision several possible extensions: First, one may want extend our analysis to n-player games in order to allow for several manipulators and puppets. Allowing for several manipulator's brings the strategic aspect between rational players back into the dynamic problem. The manipulators could cooperate using repeated games strategies and take turns in making sacrifices required to manipulate puppets to their advantage. I like to offer the following line of arguments in support of this conjecture: Consider $(m + p)$-player game with m manipulators and p puppets. Since the puppets are just myopic best response robots, we consider the game as a m-player stochastic game in which the law of motion is given by the myopic best response of the p

puppets. Now we can use the folk theorem for stochastic games by Dutta (1995) to show that the m manipulators through the construction of punishment strategies can cooperate to achieve at least the Stackelberg leaders payoff (in the sense of the Stackelberg outcome with many leaders that appeared in Bulavzky and Kalashnikov, ZAMM 1996). In such an outcome, Stackelberg leaders play Nash equilibrium among themselves given the best response output of Stackelberg followers. Manipulators should be able to do even better by reaching joint payoff maximizing outcomes given the best response of puppets. At present, I do not know whether there is an optimal manipulation strategy of manipulators in the $(m + p)$-player extension of our Cournot duopoly that is also cyclic or whether optimal manipulation strategies are monotone when games have increasing or decreasing differences everywhere. This is left for future research.

Second, we can envision extensions to other adaptive learning heuristics such as fictitious play[15], reinforcement learning, imitation, trail & error learning, etc. Third, we assumed that the manipulator knows that the puppet plays myopic best response but it is more realistic to assume that such knowledge is missing. Could the manipulator learn the learning theory of the opponent (and the nature of the noise if any)? These extensions are left for future research as well.

A Proofs and Auxiliary Results

Proof of Lemma 1 If p is upper semicontinuous in y on Y, then by the Weierstrass Theorem an argmax exist. By the Theorem of the Maximum (Berge, 1963), the argmax correspondence is upper hemicontinuous and compact-valued in x. Since p is strictly quasi-concave, the argmax is unique. Hence the upper hemicontinuous best response correspondence is a continuous best response function. Since m is upper semicontinuous and b is continuous, we have that $\hat{m}$ is upper semicontinuous. □

Proof of Lemma 2 Under the conditions of the Lemma we have by Lemma 1 that $\hat{m}$ is upper semicontinuous on $X \times X$. By the Theorem of the Maximum (Berge, 1963), M_1 is upper semicontinuous on X. If M_{n-1} is upper semicontinuous on X and $\hat{m}$ is upper semicontinuous on $X \times X$, then since $\delta \geq 0$, $\hat{m}(x', x) + \delta M_{n-1}(x')$ is upper semicontinuous in x' on X. Again, by the Theorem of the Maximum, M_n is upper semicontinuous on X. Thus by induction M_n is upper semicontinuous on X for any n.

Let L be an operator on the space of bounded upper semicontinuous functions on X

[15]One reviewer suggested that if the puppet uses fictitious play rather than myopic best response, then it is much more difficult to manipulate with a cycle. Fictitious play is an uncoupled learning heuristic. Moreover, in our Cournot example, the Stackelberg outcome is unique. Thus, it follows from Schipper (2017) that the payoff to the dynamic optimizer would be strictly above Nash equilibrium. So fictitious play can be exploited by a patient dynamic optimizers in our Cournot example although the strategy may not be cyclic. At present, the form of the optimal manipulation strategy against a fictitious player is not clear to us and is left for future research.

defined by $LM_\infty(x) = sup_{x' \in X_x}\{\hat{m}(x', x) + \delta M_\infty(x')\}$. This function is upper semicontinuous by the Theorem of the Maximum. Hence L maps bounded upper semicontinuous functions to bounded upper semicontinuous functions. L is a contraction mapping by Blackwell's sufficiency conditions (Stokey, Lucas, and Prescott, 1989). Since the space of bounded upper semicontinuous functions is a complete subset of the complete metric space of bounded functions with the sup distance, it follows from the Contraction Mapping Theorem that L has a unique fixed point M_∞ which is upper semicontinuous on X. □

Proof of Lemma 3 We state the proof just for one case. The proof of the other cases follow analogously.

(i) If p has strongly decreasing differences in (y, x) on $Y \times X$, then by Topkis (1998) b is strictly decreasing in x on X. If m has strongly decreasing differences in (x, y) on $X \times Y$, $\hat{m}(\cdot, \cdot) = m(\cdot, b(\cdot))$ must have strongly increasing differences on $X \times X$.

(ii) If p has decreasing differences in (y, x) on $Y \times X$, then by Topkis (1998) b is decreasing in x on X. Hence, if m has negative externalities, $\hat{m}(x', x) = m(x', b(x))$ must be increasing in x. □

Proof of Proposition 1 (ii) The proofs of the first two lines in the table of Proposition 1 (ii) follow directly from previous Lemmata and Amir (1996a, Theorem 2 (i)). The last two lines require a proof.

Line 3 (resp. Line 4): If m has positive externalities, and both m and p have decreasing differences (resp. m has negative externalities, and both m and p have increasing differences), and X_y is contracting, then $\bar{s}_{n+1} \leq \bar{s}_n$ and $\underline{s}_{n+1} \leq \underline{s}_n$.

We first show that in this case $M_n(x)$ has decreasing differences in (n, x) on $\mathbb{N} \times X$. We proceed by induction by showing that for $x'' \geq x'$ and for all $n \in \mathbb{N}$,

$$M_n(x'') - M_n(x') \leq M_{n-1}(x'') - M_{n-1}(x'). \tag{6}$$

For $n = 1$, inequality (6) reduces to $M_1(x'') \leq M_1(x')$ since $M_0 \equiv 0$. Since m has positive externalities and p has decreasing differences (resp. m has negative externalities and p has increasing differences), and X_y is contracting, we have by Lemma 4, line 3 (resp. line 4), that M_n is decreasing on X. Hence, the claim follows for $n = 1$.

Next, suppose that inequality (6) holds for all $n \in \{1, 2, ..., k-1\}$. We have to show that it holds for $k = n$. Consider the maximand in equation (4), i.e.,

$$\hat{m}(z, x) + \delta M_{k-1}(z).$$

Since both m and p have decreasing differences (resp. both m and p have increasing differences), we have by Lemma 3 (i), line 2 (resp. line 1), that $\hat{m}(z, x)$ has increasing differences in (z, x). $M_n(z)$ has decreasing differences in (n, z) on $\{1, 2, ..., k-1\} \times X$ by the induction hypothesis. Hence $M_n(z)$ has increasing differences in $(-n, x)$ on $\{-(k-1), ..., -2, -1\} \times X$. We conclude that the maximand is supermodular in $(z, x, -n)$

on $X_y \times X \times \{-(k-1), ..., -2, -1\}$.[16] By Topkis's (1998, Theorem 2.7.6), $M_n(x)$ has increasing differences in $(x, -n)$ on $X \times \{-k, -(k-1), ..., -2, -1\}$. Thus it has decreasing differences in (x, n) on $X \times \{1, 2, ..., k\}$. This proves the claim that $M_n(x)$ has decreasing differences in (n, x) on $\mathbb{N} \times X$.

Finally, the dual result for decreasing differences to Topkis (1998, Theorem 2.8.3 (a)) implies that both $\bar{s}_{n+1} \leq \bar{s}_n$ and $\underline{s}_{n+1} \leq \underline{s}_n$. This completes the proof of line 3 (resp. line 4) in Proposition 1 (ii). □

Auxiliary Result to Proposition 1 (ii) Proposition 1 (ii) makes no mentioning of four other cases in which the monotone differences of m and p may differ. The following proposition show that analogous results for those cases can not be obtained.

Proposition 3 *(i) If [m has positive externalities and decreasing differences, and p has increasing differences] or [m has negative externalities and increasing differences, and p has decreasing differences], and X_y is expanding, then $M_n(x)$ has neither increasing nor decreasing differences in (n, x) unless it is a valuation.*

(ii) If [m has positive externalities and increasing differences, and p has decreasing differences] or [m has negative externalities and decreasing differences, and p has increasing differences], and X_y is expanding, then $M_n(x)$ has neither increasing nor decreasing differences in (n, x) unless it is a valuation.

Proof. We just prove here part (i). Part (ii) follows analogously.

Suppose to the contrary that $M_n(x)$ has decreasing differences in (n, x). We want to show inductively that for $x'' \geq x'$ we have for all $n \in \mathbb{N}$ inequality (6). For $n = 1$, inequality (6) reduces to $M_1(x'') \leq M_1(x')$ since $M_0 \equiv 0$. Since either [m has positive externalities and p has increasing differences] or [m has negative externalities and p has decreasing differences], and X_y is expanding, we have by Lemma 4, line 3 (resp. line 4), that M_n is increasing on X. Hence, a contradiction unless $M_1(x'') = M_1(x')$.

Suppose now to the contrary that $M_n(x)$ has increasing differences in (n, x). We want to show inductively that for $x'' \geq x'$ we have for all $n \in \mathbb{N}$,

$$M_n(x'') - M_n(x') \geq M_{n-1}(x'') - M_{n-1}(x'). \tag{7}$$

For $n = 1$, inequality (7) reduces to $M_1(x'') \geq M_1(x')$ since $M_0 \equiv 0$. Since either [m has positive externalities and p has increasing differences] or [m has negative externalities and p has decreasing differences], and X_y is expanding, we have by Lemma 4, line 3 (resp. line 4), that M_n is increasing on X, which implies $M_1(x'') \geq M_1(x')$.

Furthermore, suppose that inequality (7) holds for all $n \in \{1, 2, ..., k-1\}$. We have to show that it holds for $k = n$. Consider the maximand in equation (4), i.e.

[16] A real-valued function f on a lattice X is supermodular on X if $f(x'' \vee x') - f(x'') \geq f(x') - f(x'' \wedge x')$ for all $x'', x' \in X$ (see Topkis, 1998, p. 43).

$\hat{m}(z,x)+\delta M_{k-1}(z)$. Since [$m$ has decreasing differences and p has increasing differences] or [m has increasing differences and p has decreasing differences], we have by Lemma 3 (i), line 3 or 4, that $\hat{m}(z,x)$ has decreasing differences in (z,x). Hence $\hat{m}(z,x)$ has increasing differences in $(z,-x)$. $M_n(z)$ has increasing differences in (n,z) on $\{1,2,...,k-1\}\times X$ by the induction hypothesis. We conclude that the maximand is supermodular in $(z,-x,n)$ on $X_y \times X \times \{1,2,...,k-1\}$. By Topkis's (1998, Theorem 2.7.6), $M_n(x)$ has increasing differences in $(-x,n)$ on $X \times \{1,2,...,k-1\}$. Thus it has decreasing differences in (x,n) on $X \times \{1,2,...,k\}$, a contradiction unless it is a valuation. □

Proof of Proposition 1 (iii) The proof is essentially analogous to the proof of Theorem 2 (ii) in Amir (1996a). We explicitly state where we require that $\hat{m}$ is increasing on X and X_y is expanding.

We show by induction on n that $M_n(x,\delta)$ has increasing differences in $(x,\delta)\in X \times (0,1)$. For $n=1$, the claim holds trivially since M_1 is independent of δ.

Assume that $M_{k-1}(x,\delta)$ has increasing differences in (x,δ). We need to show that $M_k(x,\delta)$ has increasing differences in (x,δ) has well. We rewrite equation (4) with explicit dependence on δ and $n=k$,

$$M_k(x,\delta)=\max_{z\in X_y}\{\hat{m}(z,x)-\delta M_{k-1}(z,\delta)\}. \tag{8}$$

Since [both m and p have increasing differences] or [both m and p have decreasing difference], we have by Lemma 3 (i), line 1 or 2, that $\hat{m}(z,x)$ has increasing differences in (z,x). $M_{k-1}(z,\delta)$ has increasing differences in (δ,z) by the induction hypothesis. That is, for $\delta''\geq\delta'$ and $z''\geq z'$,

$$M_{k-1}(z'',\delta'')-M_{k-1}(z',\delta'')\geq M_{k-1}(z'',\delta')-M_{k-1}(z',\delta'). \tag{9}$$

Since [m has positive externalities and p has increasing differences] or [m has negative externalities and p has decreasing differences] and X_y is expanding, we have by Lemma 4, line 1 or 2, that $M_{k-1}(z,\delta)$ is increasing in z on X_y. Hence both the LHS and the RHS of inequality (9) are positive. Therefore, multiplying the LHS with δ'' and the RHS with δ' preserves the inequality. We conclude that $\delta M_{k-1}(z,\delta)$ has increasing differences in (δ,z). Hence the maximand in equation (8) is supermodular in (δ,z,x) on $(0,1)\times X_y\times X$.

By Topkis's (1998, Theorem 2.7.6), $M_n(x,\delta)$ has increasing differences in (δ,x) on $X\times(0,1)$. Finally, Topkis (1998, Theorem 2.8.3 (a)) implies that $\bar{s}_n(\cdot,\delta'')\geq\bar{s}_n(\cdot,\delta')$ and $\underline{s}_n(\cdot,\delta'')\geq\underline{s}_n(\cdot,\delta')$. This completes the proof of Proposition 1 (iii). □

Auxiliary Results to Proposition 1 (iii) Proposition 1 (ii) is silent on a number of cases:

Proposition 4 *Suppose that [m has positive externalities, m has decreasing differences, and p has increasing differences] or [m has negative externalities, m has increasing differences, and p has decreasing differences] and X_y is expanding. Then $M_n(x,\delta)$ has NOT increasing differences in (δ,x) on $(0,1)\times X$ unless it is a valuation.*

Proof. Suppose to the contrary that $M_n(x, \delta)$ has increasing differences in $(\delta, x) \in (0, 1) \times X$. For $n = 1$ the claim is trivial since M_n is independent of δ.

Assume that $M_{k-1}(x, \delta)$ has increasing differences in (x, δ). We need to show that $M_k(x, \delta)$ has increasing differences in (x, δ) has well. Consider the maximand in equation (8). Since [m has decreasing differences and p has increasing differences] or [m has increasing differences and p has decreasing difference], we have by Lemma 3 (i), line 3 or 4, that $\hat{m}(z, x)$ has decreasing differences in (z, x). Hence, it has increasing differences in $(z, -x)$. $M_{k-1}(z, \delta)$ has increasing differences in (δ, z) by the induction hypothesis so that inequality (9) holds.

Since [m has positive externalities and p has increasing differences] or [m has negative externalities and p has decreasing differences] and X_y is expanding, we have by Lemma 4, line 1 or 2, that $M_{k-1}(z, \delta)$ is increasing in z on X_y. Hence both the LHS and the RHS of inequality (9) are positive. Therefore, multiplying the LHS with δ'' and the RHS with δ' preserves the inequality. We conclude that $\delta M_{k-1}(z, \delta)$ has increasing differences in (δ, z). Hence the maximand in equation (8) is supermodular in $(\delta, z, -x)$ on $(0, 1) \times X_y \times X$.

By Topkis's (1998, Theorem 2.7.6), $M_n(x, \delta)$ has increasing differences in $(\delta, -x)$ on $X \times (0, 1)$. Hence it has decreasing differences in (δ, x), a contradiction unless it is a valuation. □

Two other cases, namely

(i) [m has positive externalities and both m and p have decreasing differences] or [m has negative externalities and both m and p have increasing differences] and X_y is contracting,

(ii) [m has positive externalities, increasing differences, and p has decreasing differences] or [m has negative externalities, decreasing differences, and p has increasing differences] and X_y is contracting,

can not be dealt with the method used to prove Proposition 1 (iii) and Proposition 4. Both cases are such that according to Lemma 4 we have that $M_n(x, \delta)$ is decreasing on X. Therefore the analogous inequality to (9) may be reversed if multiplying the LHS with δ'' and the RHS with δ'.

Proof of Proposition 2 (i) Note that the first-order condition for the maximization in equation (5) (analogously for equation (4)) is

$$\frac{\partial \hat{m}(x, s(x))}{\partial z} + \delta \frac{\partial M(s(x))}{\partial x} = 0. \tag{10}$$

Suppose that for some $x'' > x'$, $s(x'') = s(x')$. Then from equation (10) we conclude $\frac{\partial \hat{m}(x'', s(x''))}{\partial z} = \frac{\partial \hat{m}(x', s(x'))}{\partial z}$, which contradicts that $\hat{m}$ has strongly decreasing differences in (x, z). Hence, $s(x'') = s(x')$ is not possible, and then by Proposition 1 (i), $s(x'') < s(x')$.

This completes the proof of part (i). $\square$

Proof of Proposition 2 (ii)

The proof is essentially "dual" to the proof of Amir (1996a, Theorem 3 (ii)).

By Proposition 1 (ii) that $s_{n+1}(x) \geq s_n(x)$ for all $x \in X$. Suppose that for some $x_n \in X$, $s_{n+1}(x_n) = s_n(x_n)$. We will show that there exists $x' \in X$ such that $s_{n-1}(x') = s_{n-2}(x')$.

Plugging $s_{n+1}(x_n) = s_n(x_n)$ in the Euler equations corresponding to the problem given in equation (4) for $n = 2, 3, ...$

$$\frac{\partial \hat{m}(s_n(x_n), x_n)}{\partial z} + \delta \frac{\partial \hat{m}(s_{n-1}(s_n(x_n)), s_n(x_n))}{\partial x} = 0, \tag{11}$$

$$\frac{\partial \hat{m}(s_{n+1}(x_n), x_n)}{\partial z} + \delta \frac{\partial \hat{m}(s_n(s_{n+1}(x_n)), s_{n+1}(x_n))}{\partial x} = 0, \tag{12}$$

leads to

$$\frac{\partial \hat{m}(s_{n-1}(s_n(x_n)), s_n(x_n))}{\partial x} = \frac{\partial \hat{m}(s_n(s_{n+1}(x_n)), s_{n+1}(x_n))}{\partial x}.$$

Since $\hat{m}$ has strongly increasing differences by Lemma 3 (i) we must have $s_{n-1}(s_n(x_n)) = s_n(s_{n+1}(x_n))$. Hence $s_{n-1}(s_n(x_n)) = s_n(s_n(x_n))$. Set $x_{n-1} \equiv s_n(x_n)$. Thus $s_{n-1}(x_{n-1}) = s_n(x_{n-1})$. Plugging into the Euler equations,

$$\frac{\partial \hat{m}(s_{n-1}(x_{n-1}), x_{n-1})}{\partial z} + \delta \frac{\partial \hat{m}(s_{n-2}(s_{n-1}(x_{n-1})), s_{n-1}(x_{n-1}))}{\partial x} = 0, \tag{13}$$

$$\frac{\partial \hat{m}(s_n(x_{n-1}), x_{n-1})}{\partial z} + \delta \frac{\partial \hat{m}(s_{n-1}(s_n(x_{n-1})), s_n(x_{n-1}))}{\partial x} = 0, \tag{14}$$

leads to

$$\frac{\partial \hat{m}(s_{n-2}(s_{n-1}(x_{n-1})), s_{n-1}(x_{n-1}))}{\partial x} = \frac{\partial \hat{m}(s_{n-1}(s_n(x_{n-1})), s_n(x_{n-1}))}{\partial x}.$$

Since $\hat{m}$ has strongly increasing differences by Lemma 3 (i) last equation implies that $s_{n-1}(s_n(x_{n-1})) = s_{n-2}(s_{n-1}(x_{n-1})) = s_{n-2}(s_n(x_{n-1}))$. Hence there exists $x' \in X$ such that $s_{n-1}(x') = s_{n-2}(x')$.

By induction we obtain the existence of $x_2 \in X$ for which $s_1(x_2) = s_2(x_2)$. The Euler equations for the one- and two-period problems at x_2 are given by

$$\frac{\partial \hat{m}(s_1(x_2), x_2)}{\partial z} = 0, \tag{15}$$

$$\frac{\partial \hat{m}(s_2(x_2), x_2)}{\partial z} + \delta \frac{\partial \hat{m}(s_1(s_2(x_2)), s_2(x_2))}{\partial x} = 0. \tag{16}$$

Since $x_2 \in X$ is such that $s_1(x_2) = s_2(x_2)$, the Euler equations imply $\frac{\partial \hat{m}(s_1(s_2(x_2)), s_2(x_2))}{\partial x} = 0$.

Note that the conditions of line 3 or 4 in Proposition 2 (ii) imply by Lemma 3 (ii) that $\frac{\partial \hat{m}(z,x)}{\partial x} < 0$, a contradiction. $\square$

References

[1] Amir, R. (1996a). Sensitivity analysis of multisector optimal economic dynamics, Journal of Mathematical Economics 25, 123-141.

[2] Amir, R. (1996b). Cournot oligopoly and the theory of supermodular games, Games and Economic Behavior 15, 132-148.

[3] Aoyagi, M. (1996). Evolution of beliefs and the Nash equilibrium of normal form games, Journal of Economic Theory 70, 444-469.

[4] Banerjee, A. and J.W. Weibull (1995). Evolutionary selection and rational behavior, in: Kirman, A. and M. Salmon (Eds.), Learning and Rationality in Economics, Blackwell Publishers, Oxford, pp. 343-363.

[5] Benhabib, J. and K. Nishimura (1985). Competitive equilibrium cycles, Journal of Economic Theory 284–306.

[6] Berge, C. (1963). *Topological spaces*, Dover edition, 1997, Mineola, N.Y.: Dover Publications, Inc.

[7] Bertsekas, D. P. (2005). Dynamic programming and optimal control, Vol I & II, 3rd edition, Belmont, MA: Athena Scientific.

[8] Boldrin, M. and L. Montrucchio (1986). On the indeterminancy of capital accumulation paths, Journal of Economic Theory 40, 26–29.

[9] Bryant, J. (1983). A simple rational expectations Keynes-type coordination model, Quarterly Journal of Economics 98, 525–528.

[10] Bulavsky, V.A. and V.V. Kalashnikov (1996). Equilibria in generalized Cournot and Stackelberg markets, Zeitschrift für Angewandte Mathematik und Mechanik 76 (S3), 387–388.

[11] Camerer, C. F., Ho, T.-H. and J.-K. Chong (2002). Sophisticated experience-weighted attraction learning and strategic teaching in repeated games, Journal of Economic Theory 104, 137-188.

[12] Chong, J.-K., Camerer, C. F., Ho, T.-H. (2006). A learning-based model of repeated games with incomplete information, *Games and Economic Behavior* **55**, 340–371.

[13] Cournot, A. (1838). Researches into the mathematical principles of the theory of wealth, MacMillan.

[14] Dubey, P., Haimanko, O., and A. Zapechelnyuk (2006). Strategic substitutes and complements, and potential games, Games and Economic Behavior 54, 77–94.

[15] Duersch, P., Kolb, A., Oechssler, J. and B. C. Schipper (2010). Rage against the machines: How subjects learn to play against computers, Economic Theory 43, 407-430.

[16] Duersch, P., Oechssler, J. and B. C. Schipper (2012). Unbeatable imitation, Games and Economic Behavior 76, 88–96,

[17] Duersch, P., Oechssler, J. and B. C. Schipper (2014). When is Tit-For-Tat unbeatable?, International Journal of Game Theory 43, 25–36.

[18] Dutta, P.K. (1995). A folk theorem for stochastic games, Journal of Economic Theory 66, 1–32.

[19] Droste,E., Hommes, C., J. Tuinstra (2002). Endogenous fluctuations under evolutionary pressure in Cournot competition, Games and Economic Behavior 40, 232-269.

[20] Ellison, G. (1997). Learning from personal experience: One rational guy and the justification of myopia, Games and Economic Behavior 19, 180-210.

[21] Fudenberg, D., Kreps, D. M. and E. S. Maskin (1990). Repeated games with long-run short-run players, Review of Economic Studies 57, 555-573.

[22] Fudenberg, D. and D. K. Levine (1998). The Theory of Learning in Games, Cambridge, M.A.: The MIT Press.

[23] Fudenberg, D. and D. K. Levine (1994). Efficiency and observability with long-run and short-run players, Journal of Economic Theory 62, 103-135.

[24] Fudenberg, D. and D. K. Levine (1989). Reputation and equilibrium selection in games with a patient player, Econometrica 57, 759-778.

[25] Güth, W. and B. Peleg (2001). When will payoff maximization survive? An indirect evolutionary analysis, Journal of Evolutionary Economics 11, 479-499.

[26] Hart, S. and A. Mas-Colell (2013). Simple adaptive strategies: From regret-matching to uncoupled dynamics, World Scientific Publishing.

[27] Hart, S. and A. Mas-Colell (2006). Stochastic uncoupled dynamics and Nash equilibrium, Games and Economic Behavior 57, 286–303.

[28] Hehenkamp, B. and O. Kaarbøe, (2006). Imitators and optimizers in a changing environment, Journal of Economic Dynamics and Control 32, 1357-1380.

[29] Heifetz, A., Shannon, C., and Y. Spiegel (2007). What to maximize if you must, Journal of Economic Theory 133, 31–57.

[30] Juang W. T., (2002). Rule evolution and equilibrium selection, Games and Economic Behavior 39, 71-90.

[31] Kordonis, I., Charalampidis, A.C., and G.P. Papavassilopoulos (2017). Pretending in dynamic games: Alternative outcomes and application to electricity markets, Dynamic Games and Applications, forthcoming.

[32] Kukushkin, N.S. (2004). Best response dynamics in finite games with additive aggregation, Games and Economic Behavior 48, 94–110.

[33] Milgrom, P. and J. Roberts (1990). Rationalizability, learning, and equilibrium in games with strategic complementarities, Econometrica 58, 1255-1277.

[34] Milgrom, P. and C. Shannon (1994). Monotone comparative statics, Econometrica 62, 157-180.

[35] Monderer, D. and L.S. Shapley (1996). Potential games, Games and Economic Behavior 14, 124–143.

[36] Osborne, M. (2004). An introduction to game theory, Oxford University Press.

[37] Puterman, M. L. (1994). Markov decision processes. Discrete stochastic dynamic programming, New York: John Wiley & Sons, Inc.

[38] Rand, D. (1978). Excotic phenomena in games and duopoly models, Journal of Mathematical Economics 5, 173–184.

[39] Schipper, B. C. (2017). Strategic teaching and learning in games, The University of California, Davis.

[40] Schipper, B.C. (2009). Imitators and optimizers in Cournot oligopoly, Journal of Economic Dynamics and Control 33, 1981-1990.

[41] Stokey, N. L., Lucas, R. E. and E. C. Prescott (1989). Recursive methods in economic dynamics, Cambridge, M.A.: Harvard University Press.

[42] Topkis, D. (1978). Minimizing a submodular function on a lattice, Operations Research 26, 305-321.

[43] Topkis, D. (1998). Supermodularity and complementarity, Princeton: Princeton University Press.

[44] Van Huyck, J., Battalio, R., and R. Beil (1990). Tacit coordination games, strategic uncertainty and coordination failure, American Economic Review 80, 234–248.

[45] Vives, X. (1999). Oligopoly pricing. Old ideas and new tools, Cambridge, M.A.: Cambridge University Press.

[46] Walker, J.M., Gardner, R., and E. Ostrom (1990). Rent dissipation in a limited access Common-Pool resource: Experimental evidence, Journal of Environmental Economics and Management 19, 203–211.

[47] Young, P. (2013). Strategic learning and its limits, Oxford University Press.