MPRA

Munich Personal RePEc Archive

Strategic control of myopic best reply in repeated games

Schipper, Burkhard C

University of California, Davis

05. April 2011

Strategic Control of Myopic Best Reply in Repeated Games*

Burkhard C. Schipper[†]

First version: November 11, 2006
This version: April 5, 2011

Abstract

How can a rational player strategically control a myopic best reply 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.

†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

How to strategically interact with others? Answers to this question are given by non-cooperative solution concepts that presume symmetry in the rationality of players. This symmetry is justified for methodological reasons: We do not want to explain (trivially) ex-post differences in behavior with an assumption of ex-ante differences among the players. Yet, in real-life situations we may be (over)confident in our ability to outwit others. This is perhaps even more apparent in the increasing interaction between humans and computers, which obviously involves an asymmetry in the rationality of players. For example, calling computers call clients to schedule appointments, some modern cars take into account slow responses by drivers and enhance the break if necessary, businesses may use programmed trading in market platforms to interact with other businesses, etc. Sometimes we view these programs as inferior to human intelligence.[1] After all computers can just do what they are programmed to do. Their response may be in some circumstances inappropriate, limited and suboptimal. Even relatively intelligent machines with learning abilities, must use some kind of learning program. Such a learning algorithm may adapt only slowly or with a lag to the situation, and is prone to strategic teaching and manipulation. Given that the opponent's rationality differs from ours, it may still not be a trivial problem to answer the question of how to interact optimally with such an opponent. In particular, how could we manipulate this opponent to our advantage? In this article we will investigate the following problem that appears to be straightforward but to our knowledge has been neglected in the literature: How can a rational player optimally control an adaptively learning opponent in a repeated strategic game?

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

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

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 reply 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 reply 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.

---

[1] For some tasks, computers perform much better than humans. E.g., a simple pocket calculator can calculate much faster than most humans the number $\sqrt[6]{123456}$. What matters for my argument is that there is interaction among players with differing levels of sophistication.

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 object function is not everywhere concave and differentiable, conditions usually required for dynamic programming (see Stokey, Lucas and Prescott, 1989). Nevertheless, it is quite natural to conjecture that the optimal strategy of the player may involve to play a (current) best reply 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 reply (see Duersch, Kolb, Oechssler and Schipper, 2010), we discovered to our surprise one subject 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.[2] This experimental discovery triggered the current analysis. Can such a cycle be optimal?

Figure 1: Cycle played by a subject

ROUND
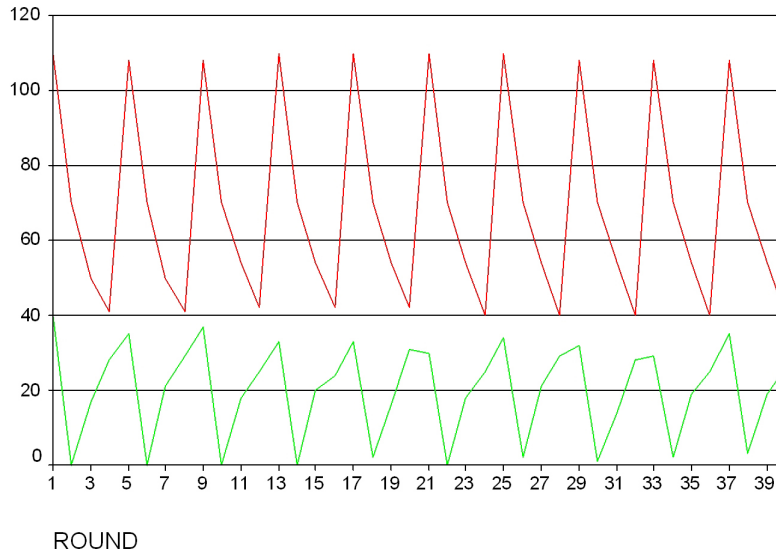
In this article we will show that if the two-player game satisfies a version of strategic substitutes or strategic complements, namely decreasing or increasing differences, then the optimal control strategy is monotone in the initial action of the opponent, the discount rate and time periods. Examples of this class of games include some Cournot duopolies

---

[2]The game was repeated over 40 rounds. The subject 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 reply with some noise. The x-axis in Figure 1 indicates the rounds, the y-axis the quantities. The lower time series depicts the computer's sequence of actions. The upper time series shows the subject's quantities. See Duersch, Kolb, Oechssler and Schipper (2010) for details of the game and the experiment.

(Amir, 1996b), Bertrand duopolies (Vives, 1999), Common pool resource games, Public goods games, Rent seeking games, Diamond 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 analogous to a Ramsey-type capital accumulation problem solved in Amir (1996a), so that his results if appropriately "translated" can be applied to our game theoretic problem. 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 can not be directly applied to our Cournot duopoly. Yet, we show in Section 3 how to use the results 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 a subject in the experiment discussed above.[3]

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 reply players who play only once. This is different from our model, in which the short-run player plays a best reply 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 reply 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, Duersch, Oechssler and Schipper, 2011a). They focus on the evolutionary selection or relative success of different boundedly rational learning rules. For instance, Droste, Hommes and Tuinstra (2002) study a population of players who can choose either a myopic best reply rule or a "Nash" rule (i.e. anticipate other Nash players and myopic best reply players) in a Cournot oligopoly. Choosing the Nash rule carries a cost. The fraction of rules used in the population is updated according to a noisy replication dynamics. They show complicated dynamics can arise. 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

---

[3]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 subject.

4

and Pareto dominant equilibrium but not vice versa. He also present 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. We provide results beyond 2x2 and 3x3 games but focus on a dynamic optimizer against a myopic best reply player.

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 A 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. $y_t \in Y_{x_t}$) the manipulator's (resp. puppet's) action, where X_y (resp. Y_x) is an upper hemi-continuous compact valued correspondence from Y to 2^X (resp. X to 2^Y). That is, we allow that a player's set of actions may depend upon the opponent'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.

Let $B : X \longrightarrow 2^Y$ be the puppet's best reply correspondence. Moreover, let the puppet's best reply function $b : X \longrightarrow Y$ be a selection of the best reply 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 reply player with a given best reply 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 reply function. We assume that $y_0 \in b(X)$, i.e., the puppet's initial action is a best reply to some action of the manipulator. We believe that this assumption is not restrictive since a best reply player should play by definition a best reply to some action of the opponent.[5]

For the existence of an optimal strategy, the manipulator's objective function should satisfy some continuity properties. While m above is assumed to be u.s.c. on $X \times Y$,

---

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

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

this property may not necessarily extend to the modified one-period objective function $\hat{m}(\cdot, \cdot) = m(\cdot, b(\cdot))$ defined on $X \times X$.

Following lemma will be useful for the study of the optimization problem of the manipulator when the puppet is a myopic best reply player.

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

The proof is contained in the appendix.

In light of Lemma 1 we will assume that m is u.s.c. on $X \times Y$ and p is u.s.c. and strictly quasi-concave on Y. Note that latter assumption is stronger than necessary. In Section 4, we discuss how to generalize it to quasiconcavity. Note that we do not impose any concavity assumption on m or $\hat{m}$.

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), 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

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

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, n runs backwards in time.

Lemma 2 *If X_y is a u.h.c. and compact-valued correspondence from Y to 2^X, m is u.s.c. on $X \times Y$, and p is u.s.c. and strictly quasi-concave in y on Y_x given $x \in X$, then for $n = 0, ...,$ the value function M_n is u.s.c. 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 in equation (4) (resp. (5))

6

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 us 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.

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 $y'', y' \in Y_{x''} \cap Y_{x'}$ 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.

A set of action $X_y \subseteq \mathbb{R}$ 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. $\max\{y'', y'\} \in F(x')$).

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. Note that according to Lemma 3 (i) whenever m and p have the same kind of monotone differences, then $\hat{m}$ has increasing differences.

Lemma 3 (Properties of $\hat{m}$)   *(i) Monotone Differences: The following table establishes relationships between increasing and decreasing differences of m, p, and $\hat{m}$:*

If m has				and p has				then m̂ has			
strongly	strictly	incr.	decr	strongly	strictly	incr.	decr.	strongly	strictly	incr.	decr.
		✓				✓				✓	
			✓				✓		✓		
	✓						✓				✓
		✓				✓					✓
	✓	✓		✓	✓	✓			✓	✓	
	✓		✓	✓	✓				✓	✓	
	✓	✓		✓	✓		✓		✓		✓
	✓		✓	✓	✓		✓		✓		✓
✓	✓	✓		✓	✓	✓		✓	✓	✓	
✓	✓		✓	✓	✓			✓	✓	✓	
✓	✓	✓		✓	✓		✓	✓	✓		✓
✓	✓		✓	✓	✓		✓	✓	✓		✓

(ii) *Monotonicity in the Second Argument: The following table 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:*

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

The proof is contained in the appendix.

With Lemma 3, some properties of n-period value functions are know from analogous results on Ramsey-type problems by Amir (1996) (see Puterman, 1994, for related results). Lemma 4 states that the n-period value functions are monotone in the previous period's action $(n+1)$ of the manipulator.

Lemma 4 *The following conclusions hold:*

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

8

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

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 $(n+1)$ of the manipulator. Proposition 1 (ii) shows that the $n+1$-horizon optimal control strategy (that gives the first period's action) is larger than the n-horizon optimal control strategy. That is, optimal control strategies are monotone over time. Finally, denote by $\bar{s}_n(\cdot, \delta)$ (resp. $\underline{s}_n(\cdot, \delta)$) be the largest (resp. lowest) optimal control strategy for the n-horizon problem when the discount rate is δ. Proposition 1 (iii) states sufficient conditions for the optimal control strategy being monotone increasing in the discount rate.

Proposition 1 *The following conclusions obtain: For $n = 1, 2, ...,$*

	If m has strictly incr. differences	m has decr. differences	*and* p has strongly incr. differences	p has decr. differences	*and* X_y is ascending	X_y is descending	*then* ... is incr. on X	... is decr. on X
(i)	✓		✓		✓		$\bar{s}_n, \underline{s}_n$	
		✓		✓	✓		$\bar{s}_n, \underline{s}_n$	
	✓			✓		✓		$\bar{s}_n, \underline{s}_n$
		✓	✓			✓		$\bar{s}_n, \underline{s}_n$
	✓	✓	✓	✓	✓		s_n	
	✓	✓	✓	✓	✓		s_n	
	✓	✓	✓	✓		✓		s_n
	✓		✓	✓		✓		s_n

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

(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 (ii) 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.[6]

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 within the cycle. Thus, apparent "irrational" behavior may in fact be rational in a dynamic context even if just finite repetitions are considered.

Example 1 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 may be finite), then it is easy to see that a cycle of $t, d, t, d, \ldots$ is optimal. If the puppet's initial action is l, such a cycle yields a payoff stream of $0, 20, 0, 20, \ldots$ whereas repeated play of the unique Nash equilibrium action $d, d, d, d, \ldots$ yields $6, 6, 6, 6, \ldots$.

Note that t is strictly dominated by d. Thus, the example demonstrates that the manipulator may use a strictly dominant action in an optimal control strategy if it induces

---

[6]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.

the puppet to a response favorable to the manipulator.[7]   □

One may also conjecture analogous results to Proposition 1 (iii) for cases in which monotone differences of m and p differ or with reverse externalities of m. We discuss this in the appendix and show some auxiliary results (Proposition 4).

Proposition 2 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).

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 $n = 1, ...,$ i.e. $s_n(x)$ is in the interior of X_x.*

(i)

If m has strongly		**and** p has strongly		**and** X_y is		**then** ... is strictly	
incr.	*decr.*	*incr.*	*decr.*	*ascending*	*descending*	*incr.*	*decr.*
differences		*differences*				*on X*	
✓		✓		✓		s_n	
	✓		✓	✓		s_n	
✓			✓		✓		s_n
	✓	✓			✓		s_n

(ii)

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

(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)).

---

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

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 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 price. To see this note that for instance $\pi(100,0) - \pi(50,0) = 800 - 2900 = -2100 < \pi(100,100) - \pi(50,100) = -100 - 50 = -150$ while $\pi(40,20) - \pi(30,20) = 1920 - 1740 = 180 > \pi(40,30) - \pi(30,30) = 1520 - 1440 = 80$.

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

$$\eta(x,y) = (108 - x - y)x.$$

This game has strongly decreasing differences everywhere and negative externalities. The graph of this payoff function is identical to the graph of the original payoff function for the range of actions $x \in [0, 109 - y]$. For this range of x the original game satisfies strictly decreasing differences. Similarly, for any n we can find the range of x_{n+1} where the smooth n-period's 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 subject in the experiment 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_{n+1} is in the range where the n-objective function coincides with the smooth n-period objective function. We show that after eight periods this assumption is violated for the fourth 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 $\Pi_n(x_{n+1})$,[8]

$$
\begin{aligned}
\Pi_1(x_2) &= \max\{109 - x_1 - b(x_2), 0\}x_1 - x_1 \\
\Pi_2(x_3) &= \max\{109 - x_2 - b(x_3), 0\}x_2 - x_2 \\
&\quad + \max\{109 - s_1(x_2) - b(x_2)\}s_1(x_2) - s_1(x_2) \\
&\quad \vdots \quad \vdots \quad \vdots
\end{aligned}
$$

and solve for the n-period optimal control strategy $s_n(x_{n+1})$ under the assumption that x_{n+1} is in the range where the n-period objective function coincides with the smooth n-period objective function:[9]

$$
\begin{aligned}
s_1(x_2) &= \frac{1}{4}x_2 + 27 \text{ if } x_2 \in [0, 108] \\
s_2(x_3) &= \frac{4}{15}x_3 + 36 \text{ if } x_3 \in [41.59, 108] \\
s_3(x_4) &= \frac{15}{56}x_4 + \frac{270}{7} \text{ if } x_4 \in [53.560, 108] \\
s_4(x_5) &= \frac{56}{209}x_5 + \frac{432}{11} \text{ if } x_5 \in [56.264, 108] \\
s_5(x_6) &= \frac{209}{780}x_6 + \frac{513}{13} \text{ if } x_6 \in [56.959, 108] \\
s_6(x_7) &= \frac{780}{2911}x_7 + \frac{1620}{41} \text{ if } x_7 \in [57.142, 108] \\
s_7(x_8) &= \frac{2911}{10864}x_8 + \frac{3834}{97} \text{ if } x_8 \in [57.191, 108] \\
s_8(x_9) &= \frac{10864}{40545}x_9 + \frac{672}{17} \text{ if } x_9 \in [57.204, 108]
\end{aligned}
$$

E.g., $s_2(x_3)$ above is the optimal two-period strategy for the original non-smooth problem if $x_3 \in [41.59, 108]$ since under the latter condition the non-smooth problem coincides with the smooth problem.

Note that if x_{n+1} 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 $s_n(x_{n+1}) = 108$ since the graph of the n-period objective function has the typical shape depicted in Figure 2.[10]

Note further that if $x_n = 108$ then $\Pi_n(x_{n+1}) = k$ for all $x_{n+1} > 1$. That is, if $x_n = 108$ then the n-periods payoff is constant in x_{n+1}. So it does not matter what the puppet

---

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

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

[10]The figure depicts as example the smooth (lower figure) and original (upper figure) n-period objective functions for $n = 2$. For $n > 2$, the graph of the objective function is qualitatively similar.

Figure 2: Objective Function for $n = 2$
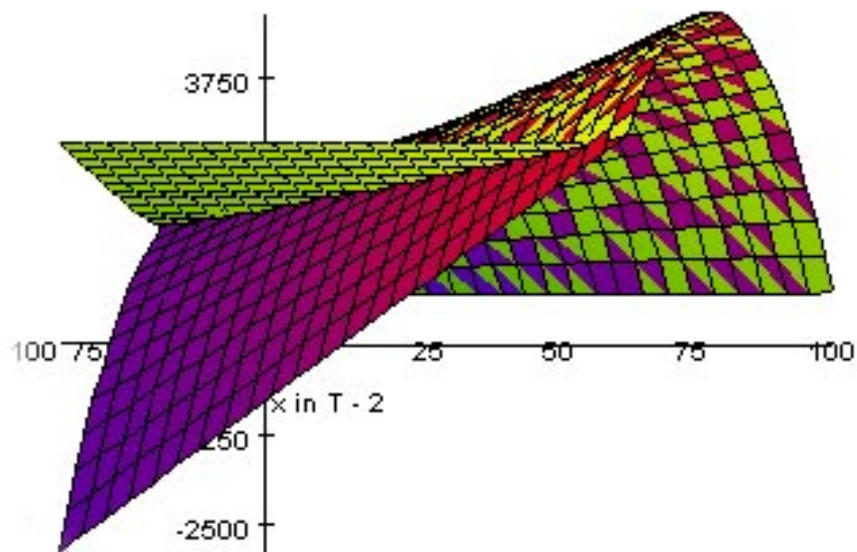
plays in n. In particular, the puppet could play a best reply to x_1, the last period's action of the manipulator. We conclude that in the n-period problem, if x_{n+1} is outside the respective range for which the n-period objective function coincides with the smooth n-period objective function, then there is an optimal cycle which starts with $x_n = 108$.

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 $x_9 = 28$ (defined by $40 = b(x_9)$) 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:

n	8	7	6	5	4	3	2	1
x_n	108	68.464	57.857	54.964	54	53.036	50.143	39.536

We note that $n = 4$ is the latest period, for which $x_{n+1} = x_5 \notin [56.264, 108]$ ($54.964 < 56.264$), a contradiction that the 8-cycle being optimal for the n-period problem. Hence a smaller cycle must be optimal. Indeed, when we compute all smaller cycles using n-

14

period optimal control strategies s_n and starting values 108, then we find that the 4-cycle is optimal.

Consider now the strategic control problems for this game with $n > 8$. Suppose that a 4-cycle is not optimal anymore for such problem with period's larger than 8. Then we must have that x_5 in optimal path for the $n > 8$ problem is strictly lower than x_5 for the 8-cycle. Otherwise, by previous arguments the 4-cycle would be optimal. This could only be true if x_8 in the optimal path of $n > 8$ period problem is strictly larger than x_8 in the 8-cycle, since by Proposition 1 (i) for $n = 1, 2, ...$ we have that s_n is monotone increasing in x_{n+1}. However, already for the 8-cycle we have $x_8 = 108$, the largest undominated action that makes the puppet leave the market in the following period. Hence, x_8 in the optimal path for the $n > 8$ period problem can not be larger, which implies that for $n = 5$ we must have that $x_5 \notin [56.264, 108]$ ($54.964 < 56.264$), a contradiction to the assertion the 4-cycle is not optimal. This completes the proof that 4-cycles are 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 $4m$ for $m = 1, 2, ...$ period. For any previous periods, there is an optimal path monotone over periods since the range-assumption won't be 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.

Finally, note that the example is not non-generic. That is, small perturbations of the payoff function π 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 reply 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. It is unlikely that using similar methods our results can be extended to the weaker ordinal notion of (dual) single crossing property. 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 u.s.c. and b continuous, since if b is just u.s.c. the composition $\hat{m}$ may not be a u.s.c. function. E.g., if b is a u.s.c. function then $-b$ is a l.s.c. function. Hence $m(\cdot, -b(\cdot))$ may not be a u.s.c. 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 l.h.c. correspondence. But the Theorem of the Maximum just yields a u.h.c. correspondence. As 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 u.h.c. correspondence, a convex-valued u.h.c. correspondence can be approximated by a closed and convex-valued l.h.c. correspondence. Note that convex-valuedness of B requires quasi-concavity of p anyway.

In our model we required the initial action of the puppet to be a best reply 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 reply players are programmed to best replies. So no matter what they play, it should be a best reply to some of the opponent's action. This is intuitive especially if we view period 0 not as the first period.

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 reply 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 dominant actions while in those zero-sum games the manipulator always plays a best reply and hence he does not need to sacrifice short term for long term gain.[11]

Any optimal cycles are due to the "mechanistic" nature of myopic best reply. 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 reply player to recognize cycles.

We view our analysis as a first step towards studying strategic control of adaptive learning. We envision several possible extensions. First, one may want extend our anal-

---

[11] 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 (2011b).

ysis to n-player games in order to allow for several manipulators and puppets. Allowing for several puppets who play myopic best reply increases the complexity of the analysis. Consider a problem with two puppets and a manipulator. A puppet plays not only a best reply to the manipulator's action to the previous period but also a best reply to any other puppet in the previous period who himself plays a best reply to the manipulator's actions in the previous-pervious period and ... Hence, the manipulator t-period objective function does not only depend on her previous period's quantity but on her entire sequence of actions up to t. Allowing for several manipulator's brings the strategic aspect between rational players back into the dynamic problem. On the other hand, if we allow several manipulators, then they could cooperate using repeated games strategies and take turns in making sacrifices required to manipulate the puppet. Second, myopic best reply is just one adaptive learning theory. Our analysis should be extended to other (adaptive) learning theories such as fictitious play, reinforcement learning, imitation trail & error learning, etc. or better to entire classes of (adaptive) learning theories. Duersch, Oechssler and Schipper (2011a) show that no strategy (incl. no dynamic optimizer) can manipulate a player following the decision rule "imitate-if-better" to his advantage in many textbook examples of games such as Cournot duopolies, Bertrand duopolies, rent seeking, common pool resource games, minimum effort coordination games, etc. Third, we assumed that the manipulator knows that the puppet plays myopic best reply but in reality such knowledge may be missing. Could the manipulator learn the learning theory of the opponent (and the nature of the noise if any)? These extensions are left for further research.

A   Proofs and Auxiliary Results

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

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

Let L be an operator on the space of bounded u.s.c. functions on X defined by $LM_\infty(x) = sup_{x' \in X_x}\{\hat{m}(x', x) + \delta M_\infty(x')\}$. This function is u.s.c. by the Theorem of the Maximum. Hence L maps bounded u.s.c. functions to bounded u.s.c. functions. T is a contraction mapping by Blackwell's sufficiency conditions (Stokey, Lucas, and Prescott, 1989). Since the space of bounded u.s.c. 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 u.s.c. 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. Since 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 $X_y \times X \times \{-(k-1), ..., -2, -1\}$.[12] 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

---

[12] 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).

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 $n \in \mathbb{N}$,

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

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

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). $\square$

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. $\qquad \square$

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). $\qquad \square$

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 $n = 2, 3, \ldots$

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

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

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

[6] 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.

[7] 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.

[8] Duersch, P., Kolb, A., Oechssler, J. and B. C. Schipper (2006). Rage against the machines: How subjects learn to play against computers, *Economic Theory* **43**, 407-430.

[9] Duersch, P., Oechssler, J. and B. C. Schipper (2011a). Unbeatable imitation, mimeo., University of California, Davis.

[10] Duersch, P., Oechssler, J. and B. C. Schipper (2011b). Pure strategy equilibrium in symmetric two-player zero-sum games, mimeo., University of California, Davis.

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

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

[13] 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.

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

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

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

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

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

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

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

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

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

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

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

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

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