Strategic Control of Myopic Best Reply in Repeated Games

Burkhard C. Schipper*
Department of Economics
University of California, Davis

Very Preliminary and Incomplete: November 11, 2006

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 and across time periods. As an interesting example outside this class of games we present a Cournot duopoly with non-negative prices and show that in a finite repetition the optimal control strategy involves a cycle.

Keywords: Strategic teaching, learning, dynamic optimization, strategic substitutes, strategic complements.

JEL-Classifications: C70, C72.

*Department of Economics, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA, Phone: +1-530-752 6142, Fax: +1-530-752 9382, Email: bcschipper@ucdavis.com

1 Introduction

How to strategically interact with others? Answers to this question are given (often dissatisfactory) 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 are often (over)confident in our ability to outwit others. On top of it, more and more we interact with computer programs, which obviously involve an asymmetry in the rationality of the players. For example, calling computers call clients to schedule appointments, businesses may use programmed trading in electronic market platforms etc. Often we view these programs as inferior to human intelligence. After all computers can just do what they are programmed to do. Their response may be inappropriate, limited and suboptimal. Even relatively intelligent machines who are able to learn, must use some kind of learning programm. Such 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 with such opponent optimally. In particular, how could we manipulate him to our advantage? In this article we will investigate such problem that appears to be straight forward but is to our knowledge 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

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

in which $x_t \in \mathbb{R}_+$ (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 previous period's quantity of the player,

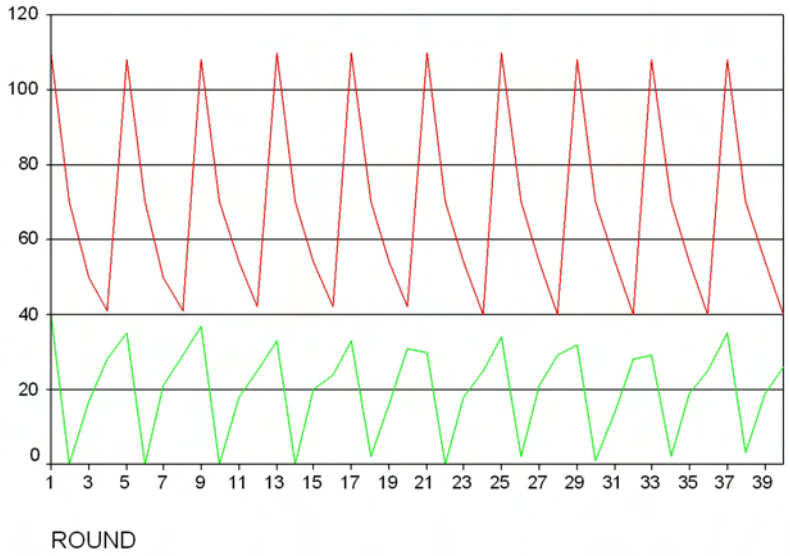
$$y_{t+1} = \max\left\{\frac{108 - x_t}{2}, 0\right\}.$$

What is the player's optimal strategy against the opponent? Is there a possibility to strategically manipulate the opponent such that he plays favorable to the player? Of course, this may require that the player forgoes some short-run profit in order to gain more in the long run.

Note that this is a dynamic programming problem that is non-standard in the sense as the object function is not everywhere concave and differentiable, conditions usually required for dynamic programming (see Stokey, Lucas and Prescott, 1989). Nevertheless we initially conjectured that the optimal strategy may involve 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 Dürsch, Kolb, Oechssler and Schipper, 2006), we discovered to

our surprise that the subject who played the following 4-cycle of quantities (upper line in Figure 1) obtained a much larger 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 subject



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 and over time periods. Examples of this class of games include some Cournot duopolies, Bertrand duopolies, Common pool resource games, Rent seeking games, Arms race etc. The key idea is to apply methods from lattice programming (Topkis, 1978, 1998) to dynamic programming (see Topkis, 1978, Puterman, 1994, Amir, 1995). It turns out that our problem is analogous to a Ramsey-type capital accumulation problem solved in Amir (1995), 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). Yet, we show how to use our general results in order 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

[1]The session is 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. The Stackelberg leader quantity is 54, the follow quantity is 27 (profit 728), the Cournot Nash equilibrium quantity 36 (payoff 1296). The computer is programmed to myopic best reply with some noise. The lower line in Figure 1 depicts the computer's sequence of actions. See Dürsch, Kolb, Oechssler and Schipper (2006) for details of the game and the experiment.

an experiment discussed above.[2]

Our approach in this paper bears some resemblance with the literature on long-run and short-run players (sometimes referred to also as long-lived and short-lived players) in infinitely repeated games (see Fudenberg, Kreps and Maskin, 1990, Fudenberg and Levine, 1989, 1994). In this literature a long-run optimizer faces a sequence of static 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. Nevertheless, this literature on reputation in repeated games is close in spirit since it recognizes how players attempt to manipulate the opponents' learning process and try to "teach" them how to play. 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 (Schipper, 2006, Hehenkamp and Kaarbøe (2006), Matros, 2004, Gale and Rosenthal, 1999). They focus on the evolutionary selection or relative success of differing boundedly rational learning rules.

The next section presents the general results. In Section 3 we discuss the cyclic example. We conclude with a discussion in Section 4. Proofs are relegated to the appendix.

2 General Results

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, a player's set of actions may depend on the opponent's action.[3]

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 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, define the puppet's best reply function $b : X \longrightarrow Y$ as a selection of the best reply correspondence, i.e., $b(x) \in B(x)$ for any $x \in X$.

[2]In fact, the average payoff of the optimal cycle is 1522, only a minor improvement over the average payoff of the subject's cycle (1520).

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

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 u.h.c. and compact valued, 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 $\hat{m} := m(\cdot, b(\cdot))$ is u.s.c. on $X \times X$ and $X_x := X_{b(x)}$ is u.h.c. and compact valued.*

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 probably stronger than necessary (see the discussion in Section 4). Note that we do not impose any concavity assumption on $\hat{m}$.

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, his action at t is

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

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

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

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

s.t. $x_{-1} \in X$ defined by $y_0 = b(x_1)$ given y_0, $x_t \in X_{x_{t-1}}$ for $t = 0, 1, ..., T-1$, and $0 < \delta < 1$.[4]

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

$$M_n(x) = \sup_{x' \in X_x} \{\hat{m}(x', x) + \delta M_{n-1}(x')\} \tag{2}$$

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

$$M_\infty(x) = \sup_{x' \in X_x} \{\hat{m}(x', x) + \delta M_\infty(x')\}. \tag{3}$$

Lemma 2 *If X_y is u.h.c. and compact valued, 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*

[4]In the Section 4 we discuss the assumption of requiring y_0 to be a response to some manipulator's quantity.

The proof is contained in the appendix.

In light of Lemma 2, optimal control strategies exist. We can replace the sup in equation (2) and (3) by max. Let $G_n(x)$ be the set of arg max in equation (2) (resp. (3)) if n is finite (resp. infinite). $G_n(x)$ is the set of all optimal decisions in the first period when the problem's horizon consists of n periods. Let g_n be a selection of G_n, and $\bar{g}_n$ and $\underline{g}_n$ be the maximum and minimum selection of G_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 $d_i : X \longrightarrow X$ and $d_i(x) \in X_x$. 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.

Before we can study properties of the solution for our dynamic optimization problem, we need to state some definitions and preliminary results. The first definitions 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 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 on Y, f is of class C^1, 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 $X_{y''} \supseteq (\supseteq) X_{y'}$. 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 objective function $\hat{m}$. These properties will allow us later on to show properties of optimal control strategies. Note that by (i) whenever m and p have the same monotone differences, then $\hat{m}$ has increasing differences.

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

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

(ii) The following table establishes a relationship 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 externalities	*negative externalities*	*increasing differences*	*decreasing differences*	*increasing in x_t*	*decreasing in x_t*
✓		✓		✓	
✓			✓		✓
	✓	✓			✓
	✓		✓	✓	

The proof is contained in the appendix.

With this lemma, the properties of solution to our dynamic programming problem are know from analogous results on Ramsey-type problems by Amir (1996) (see Puterman, 1994, for related results). Below we state the results adapted to our language (Propositions 1 to 3). The proofs follow directly from above lemmata and Amir (1996).

Proposition 1 *The following conclusions hold:*

(i)

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

(ii)

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

This proposition states that both n-period value functions and n-period optimal control strategies are monotone in the previous period's action $(n+1)$ of the manipulator.

The next proposition 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. Moreover, denote by $\bar{g}_n(\cdot, \delta)$ (resp. $\underline{g}_n(\cdot, \delta)$) be the largest (resp. lowest) optimal control strategy for the n-horizon problem when the discount rate is δ. Part (ii) of Proposition 2 shows that the optimal control strategy is increasing in the discount rate.

Proposition 2 *(i) If [m has positive externalities and p has increasing differences] or [m has negative externalities and p has decreasing differences] and X_y is expanding, then $\bar{g}_{n+1} \geq \bar{g}_n$ and $\underline{g}_{n+1} \geq \underline{g}_n$ for $n = 1, ...$*

(ii) If $\delta'' \geq \delta'$, $\delta'', \delta' \in (0,1)$, then $\bar{g}_n(\cdot, \delta'') \geq \bar{g}_n(\cdot, \delta')$ and $\underline{g}_n(\cdot, \delta'') \geq \underline{g}_n(\cdot, \delta')$.

Proposition 3 strengthens previous results to strict monotone optimal control strategies. This comes at the cost of assuming strongly increasing or decreasing differences (and hence differentiability of the payoff functions).

Proposition 3 *Let g_n be any interior optimal strategy for $n = 1, ...$, i.e. $g_n(x)$ is in the interior of X_x.*

(i) If both m and p have strongly increasing differences or strongly decreasing differences and X_y is ascending, then $g_n(x'') > g_n(x')$ if $x'' > x'$, $n = 1, ...$

(ii) If [m has positive externalities and strongly increasing differences, and p has strongly increasing differences] or [m has negative externalities and strongly decreasing differences, and p has decreasing differences] and X_y is expanding, then $g_{n+1}(x) > g_n(x)$ for all $x \in X$ and $n = 1, ...$

(iii) If both m and p have strongly increasing differences or strongly decreasing differences and $\delta'' > \delta'$, $\delta'', \delta' \in (0,1)$, then $g_n(\cdot, \delta'') > g_n(\cdot, \delta')$, $n = 1, ...$

We want to compare actions and payoffs resulting from optimal control strategies with Nash equilibrium strategies and payoffs in the one-shot game. Let (x_e, y_e) denote a pure strategy Nash equilibrium profile in the one-shot game with the manipulator's largest or smallest Nash equilibrium action. Denote by M_n^* the n-period discounted sum of stage-game Nash equilibrium payoffs to the manipulator corresponding to the n-period play of such Nash equilibrium (x_e, y_e).

Proposition 4 *Suppose that X_y is expanding, both m and p have decreasing differences and m has negative externalities (resp. both m and p have increasing differences and m has positive externalities). If $y_0 \leq y_e$ (resp. $y_0 \geq y_e$) then there exists a transition path*

induced by n-optimal control strategies g_n, $n = 1, ...$ *and* $y_0 = b(x_{-1})$ *such that for any element* x_n *for the transition path we have* $x_n \geq x_e$ *and for the corresponding payoffs holds* $M_n(x_{-1}) \geq M_n^*$.

The proof is contained in the appendix.

We are also interested in a comparison of actions and payoffs between the manipulator and the puppet. Such comparison makes only sense when payoff functions are symmetric. Let $P_n(x)$ denote the n-period discounted payoff to the puppet if the manipulator follows an n-period optimal control strategy. The next proposition shows that in games with decreasing differences and negative externalities (e.g. some Cournot duopoly) the manipulator is better off than the puppet if the puppet's initial action is not to large. This is remniscent of Stackelberg outcomes, where the manipulator takes a role similar to the Stackelberg leader and the puppet is the Stackelberg follower.

Proposition 5 *Suppose that payoff functions* m *and* p *are symmetric, i.e.* m *satisfies all properties of* p. *Suppose further that* p *(and thus* m*) has decreasing differences, negative externalities,* X_y *is expanding, and* b *has a slope above* -1. *If* $y_0 \leq y_e$ *then there exists a transition path induced by n-optimal control strategies* g_n, $n = 1, ...$ *and* $y_0 = b(x_{-1})$ *such that* $M_n(x_{-1}) \geq M_n^* \geq P_n(x_{-1})$.

The proof is contained in the appendix. The assumption that b has a slope above -1 is probably stronger than necessary. In symmetric games considered here it implies a unique Nash equilibrium which is symmetric.

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. Since the game does not satisfy decreasing differences everywhere, previous results are not applicable. To see this note that for instance $\pi(100, 0) - \pi(50, 0) = 800 - 2900 < \pi(100, 100) - \pi(50, 100) = -100 - 50$ while $\pi(40, 20) - \pi(30, 20) = 1920 - 1740 > \pi(40, 30) - \pi(30, 30) = 1520 - 1440$.

Consider now a "smooth" version of the game with symmetric payoff functions given below, in which we do not insist on a non-negative price:

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

This game has strongly decreasing differences and negative externalities everywhere. 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 is optimal. 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-objective function. We show that after eight periods this assumption is violated for the fourth period. We show that this means that there must be cycle if $n = 8$, and it turns out that the 4-cycle is optimal. Using our monotonicity results, we extend the result to $n > 8$.

For $n = 1, 2, ..., 8$, we write down recursively the n-objective functions $\Pi_n(x_{n+1})$,[5]

$$\begin{aligned}
\Pi_1(x_2) &= \max\{109 - x_1 - b(x_2), 0\}x_1 - x_1 && (4)\\
\Pi_2(x_3) &= \max\{109 - x_2 - b(x_3), 0\}x_2 - x_2 \\
&\quad + \max\{109 - g_1(x_2) - b(x_2)\}g_1(x_2) - g_1(x_2) && (5)\\
&\vdots \quad \vdots \quad \vdots
\end{aligned}$$

and solve for the n-optimal control strategy $g_n(x_{n+1})$ under the assumption that x_{n+1} is in the range where the n-objective function coincides with the smooth n-objective function:[6]

$$\begin{aligned}
g_1(x_2) &= \frac{1}{4}x_2 + 27 \text{ if } x_2 \in [0, 108] && (6)\\
g_2(x_3) &= \frac{4}{15}x_3 + 36 \text{ if } x_3 \in [44.41, 108] && (7)\\
g_3(x_4) &= \frac{15}{56}x_4 + \frac{270}{7} \text{ if } x_4 \in [53.560, 108] && (8)\\
g_4(x_5) &= \frac{56}{209}x_5 + \frac{432}{11} \text{ if } x_5 \in [56.264, 108] && (9)\\
g_5(x_6) &= \frac{209}{780}x_6 + \frac{513}{13} \text{ if } x_6 \in [56.959, 108] && (10)\\
g_6(x_7) &= \frac{780}{2911}x_7 + \frac{1620}{41} \text{ if } x_7 \in [57.142, 108] && (11)\\
g_7(x_8) &= \frac{2911}{10864}x_8 + \frac{3834}{97} \text{ if} && (12)\\
g_8(x_9) &= \frac{10864}{40545}x_9 + \frac{672}{17} \text{ if} && (13)
\end{aligned}$$

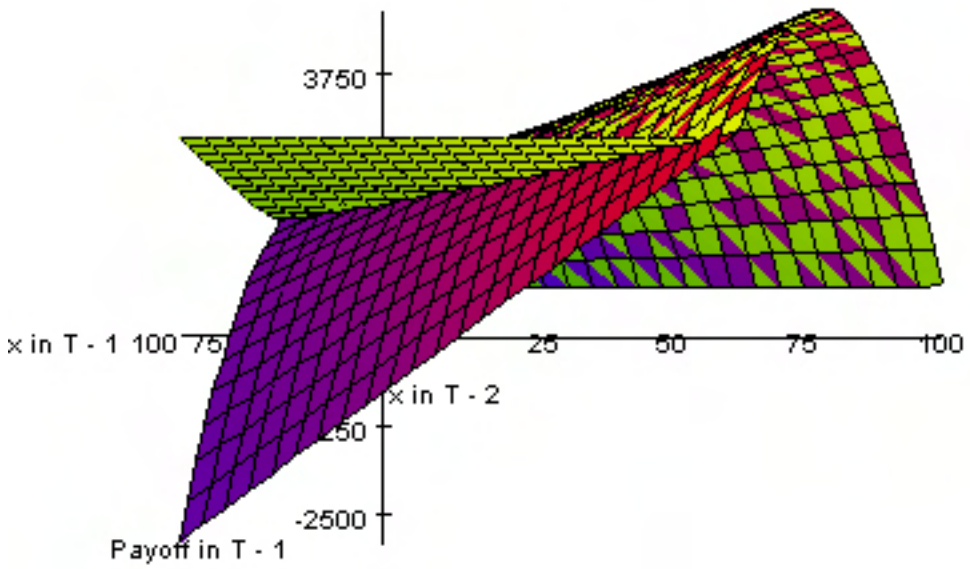
Note that if x_{n+1} is outside the respective for range for which the n-objective function coincides with the smooth n-objective function, then there is a corner solution $g_n(x_{n+1}) = 108$ since the graph of the n-objective function has the typical shape depicted in Figure 3.[7]

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

[6]Interestingly, the denominator in the linear factor in g_n is identical the nominator of the linear factor in the g_{n+1}.

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

Figure 2: Objective Function for $n = 2$



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-payoff is constant in x_{n+1}. So it does not matter what the puppet 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 for range for which the n-objective function coincides with the smooth n-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-objective function coincides with the smooth 8-objective function. Hence there must be at least an 8-cycle in the 8-period problem.

Suppose there is such 8-cycle in the 8-period problem, then by above arguments $x_8 = 108$. Using the n-optimal control strategies for $n = 1, 2, ..., 7$ above, we can compute the optimal path 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-optimal control strategies g_n and starting values 108, then we find that the 4-cycle is optimal.

Consider now any problems 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 for $n = 1, ...$ we have that g_n is monotone increasing in x_{n+1}. However, already for the 8-cycle we have $x_8 = 108$, the largest undominated action. 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 in 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 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 bounded for the price although the optimal cycle length may depend on the parameters of the game.

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. If we assume that both m and p are supermodular in actions, then $\hat{m}$ may not be supermodular even if $b(x)$ 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 property of decreasing and increasing differences to obtain our results. It is unlikely that our results extend to the weaker order notions 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 (Topkis, 1998).

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 would require that B is a convex-valued l.h.c. correspondence. The Theorem of the Maximum just yields a u.h.c. correspondence. Convex-valued-ness requires quasi-concavity of p anyway. So strict quasi-concavity of p may be weakened to quasi-concavity and a slightly stronger continuity property.

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. Why should the puppet start already with a rationalizable action? 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. However, note that for instance 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. Such 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 of studying strategic control of adaptive learning. 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. Second, myopic best reply is just one adaptive learning theory. Our analysis should be extended to other (adaptive) learning theories as well such as fictitious play, reinforcement learning, imitation, trail & error learning, etc. or better to classes of (adaptive) learning theories. Third, (adaptive) players may make mistakes. Would cycles in our Cournot example survive if the myopic best reply player would tremble to any quantity with a tiny but strict positive probability? Forth, 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)?

A Proofs

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, 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 4. Suppose that both m and p have decreasing differences and m has negative externalities. We focus on the Nash equilibrium (x_e, y_e) in which the manipulator plays her smallest Nash equilibrium action (the proof with the largest Nash equilibrium action is analogous, just replace $\underline{g}_n$ by $\bar{g}_n$). First, consider $T = 1$. Since $y_0 = b(x_{-1})$ and p has decreasing differences, we have $y_0 \leq y_e$ if and only if $x_{-1} \geq x_e$. Since m has decreasing differences we have $\underline{g}_1(x_e) = x_e$ and $\underline{g}_1(x_{-1}) \geq x_e$ for $x_{-1} \geq x_e$. This follows from Proposition 1. Consider now $T > 1$. Since m has negative externalities, p has decreasing differences and X_y is expanding, it follows from Proposition 2 that $\underline{g}_{n+1}(x_{-1}) \geq \underline{g}_n(x_{-1})$ for $n = 1, ...$ Hence we conclude $\underline{g}_{n+1}(x_{-1}) \geq x_e$ for $n = 1, ...$ if $y_0 \leq y_e$.

Suppose now to the contrary that $\max \sum_{t=0}^{T-1} \delta^t \hat{m}(x_t, x_{t-1}) < \sum_{t=0}^{T-1} \delta^t m^*$, where m^* is the manipulator's payoff from the one-shot Nash equilibrium (x_e, y_e). Since by Lemma 3, $\hat{m}(x_t, x_{t-1})$ is increasing in x_{t-1} we must have that $\hat{m}(x_e, x_{t-1}) \geq \hat{m}(x_e, x_e)$ for any $t = 0, ...$ Hence the manipulator could improve his payoff by deviating to her one-shot Nash equilibrium action. Thus we have a contradiction to $\max \sum_{t=0}^{T-1} \delta^t \hat{m}(x_t, x_{t-1})$ being the payoff from the

optimal control strategy.

The proof is analogous if both m and p have increasing differences and m has positive externalities. □

Proof of Proposition 5. Since the game is symmetric and b has a slope above -1, the stage-game Nash equilibrium is unique and symmetric (see Vives, 1999, Section 2.3.2, Remark 17). Note that by Proposition 4, if $y_0 \leq y_e$ then $M_n(x_{-1}) \geq M_n^*$ with $y_0 = b(x_{-1})$. Also by Proposition 4, if $y_0 \leq y_e$ then for any element x_n of the transition path induced by the optimal control strategies g_n we have $x_n \geq x_e$. Since p has decreasing differences and the unique Nash equilibrium is symmetric, $y_n \leq y_e$. We want to show that $p(x_e, x_e) \geq p(y_n, x_n)$. Suppose to the contrary that $p(x_e, x_e) < p(y_n, x_n)$. If $y_0 \leq y_e$ we have by Proposition 4, $x_n \geq x_e$, $n = 1, ...$ By negative externalities, $p(y_n, x_n) \leq p(y_n, x_e)$. Hence $p(x_e, x_e) < p(y_n, x_e)$, a contradiction to x_e being the Nash equilibrium action. Thus $M_n(x_{-1}) \geq M_n^* \geq P_n(x_{-1})$. □

References

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

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

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

[4] Camerer, Colin F., Ho, Teck-Hua and Juin-Kuan Chong (2002). Sophisticated experience-weighted attraction learning and strategic teaching in repeated games, *Journal of Economic Theory* **104**, 137-188.

[5] Chong, Juin-Kuan, Camerer, Colin F., Ho, Teck H. (2006). A learning-based model of repeated games with incomplete information, *Games and Economic Behavior* **55**, 340-371.

[6] Dürsch, Peter, Kolb, Albert, Oechssler, Jörg and Burkhard C. Schipper (2006). Rage against the machines: How subjects learn to play against computers, University of California, Davis.

[7] Fudenberg, Drew, Kreps, David M. and Eric S. Maskin (1990). Repeated games with long-run short-run players, *Review of Economic Studies* **57**, 555-573.

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

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

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

[11] Gale, D. and Rosenthal, R. W. (1999). Experimentation, imitation, and stochastic stability, *Journal of Economic Theory* **84**, 1-40.

[12] Hehenkamp, B. and O. Kaarbøe, 2006. Imitators and optimizers in a changing environment, University of Dortmund.

[13] Matros, Alexander (2004). Simple rules and evolutionary selection, University of Pittsburgh.

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

[15] Schipper, Burkhard C. (2006). Imitators and optimizers in Cournot oligopoly, The University of California, Davis.

[16] Stokey, Nancy, L., Lucas, Robert E. and Edward C. Prescott (1989). *Recursive methods in economic dynamics*, Cambridge, M.A.: Harvard University Press.

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

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

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