Exploiting Myopic Learning

Mohamed Mostagir

Social and Information Sciences Laboratory
California Institute of Technology
mosta@caltech.edu

Abstract. We show how a principal can exploit myopic social learning in a population of agents in order to implement social or selfish outcomes that would not be possible under the traditional fully-rational agent model. Learning in our model takes a simple form of imitation, or replicator dynamics; a class of learning dynamics that often leads the population to converge to a Nash equilibrium of the underlying game. We show that, for a large class of games, the principal can always obtain strictly better outcomes than the corresponding Nash solution and explicitly specify how such outcomes can be implemented. The methods applied are general enough to accommodate many scenarios, and powerful enough to generate predictions that allude to some empirically-observed behavior.

1 Introduction and Related Work

The assumptions imposed on the traditional rational agent can be too restrictive, requiring instantaneous reaction to changes in the environment, perfect look-ahead and planning skills, and unlimited computational resources. In reality, even if individuals are interested in maximizing their own welfare, they may be unable to do so because of a myriad of reasons. For example, it maybe the case that finding an optimal course of action is computationally difficult or even infeasible. It can also be that agents utilize a decision making process that is different from what the traditional model dictates. For instance, they may partially or wholly base their decisions on the actions of other agents rather than carefully charting out their own course. In this paper, we deal with the following question: if we relax some of the assumptions about rationality and consider agents that do not act in full compliance with the traditional agent model, can we leverage the resulting framework to implement better outcomes, either for society or for the principal designing the system?

This is a question of mechanism design, of course. Some of the concerns above have been and continue to be addressed by algorithmic mechanism design; a subfield of mechanism design that concerns itself with computability issues [11], but it is only recently that behavioral aspects have been taken into consideration in mechanism design. This is perhaps a little surprising, given the advanced state of behavioral and experimental game theory, two of the field's basic building blocks. One possible reason for this lag in development is the many ways in

A. Saberi (Ed.): WINE 2010, LNCS 6484, pp. 306–318, 2010.

which behavior can deviate from the classical agent model, making it difficult to develop an all-encompassing behavioral framework. In this paper, we take a small step in this direction by utilizing a simple form of social learning dynamics to set up a model that allows a system designer (henceforth referred to as the principal) to manipulate social learning to his advantage.

The social learning model we employ in this paper is that of replicator dynamics [3]. This class of learning dynamics was developed in an attempt to understand how a population arrives at a steady state of a dynamical system, and was further pursued in economics as an explanation to how agents arrive at a Nash equilibrium. Under this model, an infinite pool of agents plays a game repeatedly. After each round of the game, agents are paired together randomly to compare and contrast payoffs. If agent i is paired with agent j and agent j has obtained a better payoff than i in the last round of the game, then i considers switching to j's strategy in the next round with a probability that is proportional to the difference in payoffs between the two. This way the proportion of strategies that are performing better than average grows in the population as the share of poorly-performing strategies shrink, and more often than not these dynamics lead to a Nash equilibrium of the underlying game[1]. What makes replicator dynamics particularly appealing is that it is perhaps the most rudimentary form of learning dynamic that nicely straddles the line between behavioral and rational models. On one hand, agents are updating their strategies in a myopic fashion based on simple comparisons with how their peers are doing, but on the other hand this seemingly simple behavior can and does lead to fully rational equilibrium outcomes. The canonical selfish-routing model is one example amongst many where agents converge to a Nash equilibrium by following a replicator dynamic [6]. Another nice behavioral aspect captured by the model is the tendency of human decision makers to fall into habit, as a result of the aversion to try new strategies if one is unaware of others for whom these strategies have performed well. Even in the case of meeting others with more successful strategies, the switching is only probabilistic, underlying the fact that switching to a new strategy is not always costless.

The central idea developed in this paper revolves around the indirect influence that a principal can exert on agents' decisions via exploiting the learning dynamic discussed above. We will focus on games where the principal and the population's interests are diametrically opposed, though the methods readily extend to other settings as we discuss in Section 5. We will give a formal definition of the class of games we consider in Section 2.1, but an informal description follows. There is an infinite population where each member has the choice of one out of two pure actions. For simplicity, we can think about these actions as whether to cheat or to be honest. There is a multitude of examples that fall under this setting: agents can decide whether to cheat on their taxes or not, whether to break the speed limit, put low effort into their work, etc. The principal's action against each member of the population is either to audit the agent, at a cost, or to ignore

[1] For example, replicator behavior leads to equilibrium in Prisoner's dilemma, Battle of the sexes, and a large variety of coordination and routing games (see [7]).

and run the risk of incurring a higher cost if the agent is cheating. Agents are interested in maximizing their payoffs, while the principal tries to minimize his cost. The game is repeated indefinitely. Because the population is infinite, the principal's move in each round consists of choosing a fraction of the population to audit. Under the traditional rationality assumptions this game has a unique Nash equilibrium where the agents cheat with some fixed probability and the principal audits the same fraction of the population in each round. Ideally, the principal would like to do little auditing *and* have the population stick to his desired outcome of as little cheating as is reasonable within the payoff structure of the game

The primary contribution of this paper is twofold. On the conceptual front, we argue that imperfect decision making –in its various formats– can in some cases be considered a resource that the system planner should utilize. The second contribution is methodical, where we take the main idea and build a framework that implements it in the context of naive social learning. While the abstract idea behind our framework is simple, the implications can sometimes be quite surprising. One counter-intuitive outcome of the model is that the principal's optimal strategy always makes things temporarily worse for everybody, including possibly the principal himself, in order to achieve better outcomes later. Moreover, as we discuss later in the paper, the application of the model to some real-life problems result in findings that correspond to empirically-observed behavior. This suggests that the approach proposed in this paper not only provides a normative prescription for optimizing systems with a social learning component, but also describes how some existing systems actually operate.

There has been a lot of recent work on social learning and when it can lead a society of agents to converge to the true value of an underlying state of the world, the so-called 'wisdom of the crowds' effect (for example, [2], [1]). While it would be interesting to investigate whether this kind of learning is susceptible to manipulation by a principal, it is outside our scope of interest since we explicitly focus on agents in a behavioral setting, unlike the fully-rational Bayesian agents employed in the work above. Manipulating Bayesian agents, albeit outside of a social learning setting, has been the recent focus of some work [9]. Other recent work that aims to explore the boundaries of mechanism design under behavioral assumptions is auction design for *level-k* bidders [4]. In this paper, the authors show that under such an experimentally-plausible model, it is possible to obtain revenues that are higher than those generated by Myerson's optimal auction [10].

Finally, repeated games and reputation building is a topic with an extensive body of work in the economics literature. The main results in this area are folk theorems that show what outcomes can be obtained if a game is repeated indefinitely. The traditional approach to proving such results relies on retaliation and punishment among players, a method that fails in a setting with a large population, since the identity of a deviator cannot be detected [8]. Indeed, for the class of games we consider here, the unique equilibrium of the repeated game is the same as the one-shot version and no better outcomes can be implemented under the rational model.

2 Cheat-Audit Games

In this section we consider a class of 2 × 2 games that encompasses a large number of scenarios. We call this class of games Cheat-Audit games. In these games, as mentioned in Section 1, a very large population of agents plays an infinitely repeated game against a principal. In each round of the game an agent has one of two choices, a 'safe' choice with low payoffs, and a risky choice with a higher payoff. For example, in a tax-auditing situation the safe choice would be to report honestly, whereas cheating is a choice that can provide a higher payoff if the agent is not audited by the principal. The principal on the other hand faces a choice between a costly and a costless action when it comes to dealing with each agent. In the context of the preceding examples, a costly action for the tax scenario would be to audit an agent, and a costless action would be to ignore that agent. Of course, it might be the case that auditing leads to catching a cheating agent, in which case the principal obtains a higher payoff than if he had chosen the costless action. By the same token, not auditing an honest agent is a better action for the principal, since auditing an honest agent expends auditing resources with no useful returns to the principal.

	A	I
C	0,c_1	v_3,c_3
H	v_1,c_2	v_2,0

Fig. 1. The Cheat-Audit Game

2.1 The Game

To formalize the preceding discussion, the payoffs of the game are as shown in Figure 1, with the principal being the column player. Each agent is considered a row player and has the row player's payoffs. The actions available to an agent is to either be honest (action H) or cheat (action C). The principal either audits (action A) or ignores (action I) each agent. An agent's payoffs satisfy $0 < v_1 \leq v_2 < v_3$. To conserve notation, we will assume that $v_1 = v_2$, so that an agent is indifferent to auditing as long as he is honest. An agent is interested in *maximizing* his payoff, while the principal is interested in *minimizing* his cost, where the costs satisfy $0 < c_1 < c_2 < c_3$. There is thus an implicit constraint on the principal's resources, since auditing with no gain (outcome (H, A)) is more costly than auditing a guilty agent (outcome (C, A)). The principal's preferred outcome is (H, I), where no auditing cost is incurred and no crime is committed, and the payoff to this outcome is normalized to zero. Similarly, an agent's least preferred outcome is (C, A), and is also normalized to zero. Notice that the principal's least preferred outcome, (C, I), is also the agent's most preferred one. Because of the large population assumption, the principal's action consists of choosing a fraction $0 \leq \alpha \leq 1$ of the population to which he will apply action A. We will call this fraction the *audit rate*. The upper bound on α does not have to be equal to 1, but can instead be set to $\bar{\alpha}$ to indicate that it is not possible to audit the whole population.

This diametric opposition of the principal and agents' interests suggests that the game has no pure strategy equilibria, as indeed can be checked from the figure and the relationship between the various payoffs. In fact, similar to a game of matching pennies, the single stage game as well as its repeated version possess only a unique mixed equilibrium. Let the audit rate and the fraction of C players in the fully rational setting be given by α_{Nash} and x_{Nash}, respectively. It is straightforward to verify that

$$\begin{aligned} \alpha_{Nash} &= \frac{v_3 - v_2}{v_3}; \\ x_{Nash} &= \frac{c_2}{c_3 + c_2 - c_1} \end{aligned} \tag{1}$$

As mentioned, we consider this game in an infinitely-repeated setting. Each moment in time, the game in Figure 1 is played. We will let the state of the system at time t be the fraction of the population taking action C at that time, and we will denote this fraction by $x(t)$. The principal's choice of audit rate at time t is denoted by $\alpha(t)$. Given a state $x(t)$, audit rate $\alpha(t)$, and denoting the payoff to the principal at time t by $g(t)$, we can write

$$\begin{aligned} g(x(t), \alpha(t)) &= c_1\alpha(t)x(t) + c_2\alpha(t)(1 - x(t)) + c_3(1 - \alpha(t))x(t) \\ &= (c_1 - c_2 - c_3)\alpha(t)x(t) + c_2\alpha(t) + c_3x(t) \end{aligned} \tag{2}$$

where the terms in (2) correspond to the costs discussed above. The first term is the cost associated with catching offending agents, the second term represents the cost of auditing honest agents, and the last term is the cost of ignoring agents who were in fact playing action C.

2.2 Learning Dynamics

The learning dynamics work as follows. After each round of the game, members of the population are randomly matched to compare and contrast strategies and payoffs. Under our model, there are only two possible scenarios that can lead to switching strategies: an agent who obtained the outcome (C, A) considers changing his strategy if he meets an agent who played H. Similarly, an agent who played H considers changing his strategy to C if he meets an agent who obtained the outcome (C, I). The probabilities with which these changes in strategy occur depend on the differences in payoffs between agents, as well as a transmission factor $k > 0$. We will think of k as a 'speed of transmission': the willingness of an agent to change their strategy when faced with a potentially better one. Without loss of generality, we will assume that an agent obtaining payoff u switches to the strategy of an agent getting payoff v with probability $\max\{0, \frac{v-u}{v}\}$. From Figure 1, the probability of switching in the first scenario is simply $\min\{k\frac{v_1-0}{v_1} = k, 1\}$. The probability of switching in the second scenario is given by $\min\{k\frac{v_3-v_1}{v_3}, 1\}$. It is important to stress that the way these probabilities are defined does not affect any structural results we obtain. Any scheme where the switching probabilities are proportional to the payoff differences essentially leads to the same results. We will make the derivations less cumbersome and more general by assuming that the switch in scenario one happens with probability p and in scenario 2 with probability q. We can later substitute for p and q with whatever values that are appropriate for the application under consideration. Utilizing this notation, the fraction of switchers from C to H at any moment t is equal to the fraction of C players who were audited, $\alpha(t)x(t)$, multiplied by the probability of meeting an H player, which is $1 - x(t)$, times the probability of switching p. Likewise, the fraction of switchers from H to C is equal to the fraction of H players, $1 - x(t)$, who meet C players that were not audited, which is $x(t)(1 - \alpha(t))$, multiplied by the probability q. We can then write the dynamics of the system as a function of $x(t)$ and $\alpha(t)$

$$\begin{aligned}\dot{x}(t) = f(x(t), \alpha(t)) &= q(1-\alpha(t))x(t)(1-x(t)) - p\alpha(t)x(t)(1-x(t)) \\ &= x(t)(1-x(t))(q - \alpha(t)(q+p)) \qquad (3)\end{aligned}$$

2.3 Objective

The principal's problem is now the following. Given the different values in Figure 1, the parameters of the problem, and the learning dynamics, the principal is interested in minimizing his long-run discounted cost subject to those dynamics. This long-run cost is the sum of all costs accrued from playing the game over time. Recall that the payoff at time t is given by (2). The problem can then be written as

$$\begin{aligned}&\min_{\alpha(t)} \int_0^\infty e^{-rt}((c_1 - c_2 - c_3)\alpha(t)x(t) + c_2\alpha(t) + c_3x(t))dt \qquad (4)\\ &\text{s.t.} \qquad \dot{x}(t) = x(t)(1-x(t))(q - \alpha(t)(q+p)) \\ &\qquad\qquad 0 \le \alpha(t) \le 1\end{aligned}$$

where $0 \le r < 1$ is a discount factor. Thus the principal's problem involves finding the function $\alpha^*(t)$ that solves (4). Like any dynamic problem, the difficulty facing the principal is that current decisions affect not only the immediate cost but also future costs through the dependence of the rate of change of $x(t)$ on $\alpha(t)$.

3 Optimal Policy

3.1 Single Round

Before delving into finding the optimal solution to (4), let us first develop an intuition by considering the solution if the game is played only once. The stage game cost described by (2) can be factored and rewritten as

$$g(x,\alpha) = c_3 x + \alpha(c_2 + (c_1 - c_2 - c_3)x)$$

and is obviously a linear function in α. This implies that depending on the value of x, α takes the values of either 0 or 1 in the optimal solution. Specifically, the optimal solution to the single period problem is given by

$$\alpha^* = \begin{cases} 0, \, x < \frac{c_2}{c_2+c_3-c_1}; \\ 1, \, x \geq \frac{c_2}{c_2+c_3-c_1}. \end{cases} \tag{5}$$

which is well defined because of the relationship stipulated on the costs. Thus, *assuming that x is known*, the optimal solution to a single period problem takes the form of a threshold rule: if the fraction of C players is low enough, it does not pay to audit anybody since the cost of auditing honest agents outweighs the gains from catching C players. Conversely, when the concentration of C players is over a certain level, then it is always better to audit indiscriminately since the costs incurred in auditing H players are more than made up for by catching every single C player in the population. It is easy to see that the optimal cost $g^*(x)$ is a concave function of x:

$$g^*(x) = \begin{cases} c_3 x, & x < \frac{c_2}{c_2+c_3-c_1}; \\ c_2 + (c_1 - c_2)x, & x \geq \frac{c_2}{c_2+c_3-c_1}. \end{cases} \tag{6}$$

We will see that a part of the single period solution, where a crackdown occurs if the fraction of C players is above a certain threshold and nothing is done otherwise, is somewhat retained in the solution to the general problem. The nature of the optimal cost implies that, from a strictly policing viewpoint, the principal may prefer a higher ratio of cheaters in the population to a lower one, since it increases the rate of successful audits and incurs a lower overall cost than scenarios where resources are expended without additional benefit.

3.2 General Policy

We will derive the optimal policy for (4) by formulating the Hamiltonian function and using the Euler-Lagrange equation. We assume that the principal knows $x(0)$, the initial state of the system. This is without loss of generality, since if that was not the case then the large population assumption together with the law of large numbers and the fact that state transitions happen with probability 1 ensure that the principal can initially determine the state of the system by auditing a random sample of the population. The current value Hamiltonian function for the problem maps triplets $(x, \alpha, \lambda) \in [0, 1] \times [0, 1] \times R$ to real numbers and is given by

$$\begin{aligned} H(x, \alpha, \lambda) &= g(x, \alpha) + \lambda f(x, \alpha) \\ &= c_3 x + \alpha(c_2 + (c_1 - c_2 - c_3)x) + \lambda x(1 - x)(q - \alpha(q + p)) \\ &= c_3 x + \lambda q x(1 - x) + \alpha(c_2 + (c_1 - c_2 - c_3)x - \lambda(p + q)x(1 - x)) \end{aligned} \quad (7)$$

where λ is a co-state variable that one can think of as a price attached to the change induced in x through the decision α. Of course, like the state x and the control α, λ itself is also a function of time, but the power of the Hamiltonian approach is that it reduces the general problem to an essentially single period one. The following lemma utilizes the Hamiltonian to provide necessary (but not sufficient) conditions on the optimal control trajectories.

Lemma 1. *The optimal control for Problem (4) is a bang-bang solution.*

Proof. A bang-bang solution implies that $\alpha(t)$ takes on extremal values in its domain until the solution trajectory reaches a final state. Let us denote by $\alpha^*(t)$ and $x^*(t)$ the optimal control and state trajectories. By the Minimum Principle, it must hold at each moment in time that

$$\begin{aligned} \alpha^*(t) &= \underset{0 \le \alpha \le 1}{\arg\min}\ H(x^*(t), \alpha, \lambda(t)) \\ &= \underset{0 \le \alpha \le 1}{\arg\min}\ c_3 x + \lambda q x(1 - x) + \alpha(c_2 + (c_1 - c_2 - c_3)x - \lambda(p + q)x(1 - x)) \end{aligned}$$

Similar to the single period problem, the Hamiltonian is a linear function in α. Minimizing the Hamiltonian w.r.t α, we find that the optimal control trajectory, $\alpha^*(t)$ satisfies

$$\alpha^*(t) = \begin{cases} 0, & \lambda(t) < \frac{c_2 + (c_1 - c_2 - c_3)x(t)}{(p+q)x(t)(1-x(t))}; \\ 1, & \lambda(t) > \frac{c_2 + (c_1 - c_2 - c_3)x(t)}{(p+q)x(t)(1-x(t))}; \\ [0, 1], & \lambda(t) = \frac{c_2 + (c_1 - c_2 - c_3)x(t)}{(p+q)x(t)(1-x(t))}. \end{cases} \quad (8)$$

Thus α assumes values at the boundary except when $\lambda(t) = \frac{c_2 + (c_1 - c_2 - c_3)x(t)}{(p+q)x(t)(1-x(t))}$, in which case α disappears from the Hamiltonian and can be set to any value in its domain. However, as we will see shortly, on the optimal control and state trajectories this case cannot happen except for precisely a single pair (α^*, x^*).

Lemma 1 implies that, except for the third case where the co-state variable is exactly equal to the R.H.S, the optimal control either audits the whole population or does nothing. This provides some information about the structure of the optimal policy, but not enough to completely characterize it. To do this, let us formulate (4) as a calculus of variations problem. From (3), we have

$$\alpha(t) = \frac{1}{p + q}\left(p - \frac{\dot{x}(t)}{x(t)(1 - x(t))}\right)$$

Substituting this into the objective, the problem becomes

$$\min_{x(t)} \int_0^\infty e^{-rt} g\left(x(t), \frac{1}{p+q}\left(p - \frac{\dot{x}(t)}{x(t)(1-x(t))}\right)\right) dt$$

$$= \min_{x(t)} \int_0^\infty e^{-rt} \left(c_3 x(t) + \frac{(c_2 + (c_1 - c_2 - c_3)x(t))\left(p - \frac{\dot{x}(t)}{(1-x(t))x(t)}\right)}{p+q} \right) dt \quad (9)$$

The solution to (9) provides a necessary condition on the optimal state trajectory. Specifically, the following lemma tells us that there is a constant for which the integral in (9) is stationary.

Lemma 2. *Let $x^*(t)$ be the minimizer to (9), then $x^*(t) = C$, where C is a constant that depends on the parameters of the problem.*

Proof. See Appendix.

We now fully characterize the optimal policy.

Theorem 1. *There is a value $\bar{x}$ such that the optimal policy audits everybody whenever $x(t) > \bar{x}$ and does nothing when $x(t) < \bar{x}$. If $x(t) = \bar{x}$ then the optimal policy sets $\alpha^*(t) = \frac{q}{p+q}$ and the system stays in this state indefinitely.*

Proof. We will show that the policy in the statement of the theorem is optimal by showing that an optimal policy exists and that only the policy given in the theorem satisfies the necessary conditions for an optimum. That an optimal policy exists follows from the boundedness of the cost per stage and the continuity of both g and f in the compact sets $x(t)$ and $\alpha(t)$. The presence of the discount factor ensures that the value of the optimal solution is $< \infty$.

From Lemma 2, we know that a necessary condition for the optimal path $x^*(t)$ to minimize (9) (and consequently, (4)), is that $x^*(t)$ is a constant, which we will denote by $\bar{x}$ (where $\bar{x}$ is as given in the proof of Lemma 2). This implies that as soon as $x^*(t) = \bar{x}$ there should be no further changes in the system, so that $\dot{x}^*(t)$ is equal to zero. Given the system dynamics in (3), this occurs if

$$\begin{aligned} f(x^*(t), \alpha^*(t)) &= 0 \\ x^*(t)(1 - x^*(t))(q - \alpha^*(t)(q+p)) &= 0 \end{aligned}$$

For any nontrivial specification of the problem, $\bar{x}$ is neither equal to zero or one, and hence the only solution to the above equation is $\alpha^*(t) = \frac{q}{p+q}$. From (8), we have to have $\lambda(t) = \frac{c_2 + (c_1 - c_2 - c_3)\bar{x}}{(p+q)\bar{x}(1-\bar{x})}$. The R.H.S of this is a constant, and hence $\dot{\lambda}(t) = 0$ and the system remains in the state $(\bar{x}, \frac{q}{q+p})$ forever.

Now consider any trajectory that sets $\alpha(t) \neq 1$ when $x^*(t) > \bar{x}$. By Lemma 1, if $x^*(t) \neq \bar{x}$ and $\alpha(t) \neq 1$ then $\alpha(t) = 0$[2] , in which case $\dot{x}(t) > 0$ and $x(t)$ increases. Let $x(t_1) > \bar{x}$ and $\alpha(t_1) = 0$, then for $t_2 > t_1$, $x(t_2) > x(t_1)$, i.e. the

[2] The Minimum Principle posits the following condition on $\dot{\lambda}(t)$; $\dot{\lambda}(t) = -\frac{\partial H(x^*(t), \alpha^*(t), \lambda(t))}{\partial x}$, so that the third case in (8) cannot hold unless $x(t)$ is a constant.

system moves farther from $\bar{x}$. However, because of Lemma 2, we know that the system should eventually move *towards* $\bar{x}$. Since the system is continuous, the trajectory going from $x(t_2)$ to $\bar{x}$ has to pass through $x(t_1)$ again, at which point the system returns to the same state it was in at time t_1, but with the additional cost accrued between times t_1 and t_2 added to the total cost, indicating that such a scenario cannot be optimal, and that it would have been cheaper to set $\alpha(t_1) = 1$. The reverse argument applies in the case of $x(t) < \bar{x}$.

Thus the optimal policy drives the fraction $x(t)$ to its steady state value as quickly as possible, by not doing anything when $x(t) < \bar{x}$ or by cracking down on the population when $x(t) > \bar{x}$. Once the steady state is reached, the system stays there forever through fixing the audit rate at the value given in the statement of the theorem.

4 Discussion

4.1 Comparison with Nash Equilibrium

It is natural to ask how the behavioral solution for the class of games we considered fares in comparison to the fully rational Nash equilibrium outcome. We have already discussed in Section 2.1 that the (fully rational) repeated game possesses a unique equilibrium, given by (1). This equilibrium is also a *center* of the repeated behavioral game. This means that, under the replicator assumption, the principal has a strategy such that if the game is played long enough, the fraction with which each action is played is the same as the corresponding fraction in the Nash equilibrium [7], i.e. the principal can implement the Nash outcome in the behavioral setting, if he so desires. However, the optimal solution that we obtained in this paper is not the Nash equilibrium, indicating that the Nash solution is dominated by the policy in Theorem 1. Furthermore, as soon as the game reaches steady state, the optimal policy does *less* auditing than the Nash solution. Let us denote the audit rate in the behavioral setting by α_B. From Theorem 1, α_B is given by $\frac{q}{p+q}$. Replacing p and q by the values from Section 2.2, we have $p = k$ and $q = k\frac{v_3 - v_1}{v_3}$, and hence

$$\alpha_B = \frac{q}{p+q} = \frac{\frac{v_3 - v_1}{v_3}}{1 + \frac{v_3 - v_1}{v_3}} \tag{10}$$

which is always *strictly less* than the Nash audit rate in (1). Because of this, the Nash solution never coincides with the policy in Theorem 1, so that the optimal solution always gives a strictly better outcome for the principal while at the same time reducing the amount of auditing required. It is worth noting that the speed of transmission k has no effect on α_B.

4.2 An Empirical Example

We have analyzed our model in a continuous time framework. In reality however, many of the games that fit the model take place in discrete time, or the resources required by the optimal solution can be infeasible to implement forever. In both of these scenarios, the level of $x(t)$ inadvertently increases above $\bar{x}$, and hence the optimal solution cracks down on the population by setting α^* to its maximum possible value, in an attempt to bring $x(t)$ back to $\bar{x}$. Because of the discreteness, the crackdowns always bring the value of $x(t)$ below $\bar{x}$, hence leading to a short period of low activity on the principal's part. The whole cycle is then repeated as $x(t)$ increases again. These periodic crackdowns are widely observed in many situations. In a recent paper [5], the authors empirically observe crackdowns by the police on speeders in Belgium. The paper mentions the periodicity of such crackdowns, but does not provide an explaination for such behavior. It is also mentioned that crackdowns are planned as early as a month in advance. Both of these observations are explained by our model. The recurrence of the crackdowns takes place as the police tries to bring the fraction of speeders to an optimal level, and since the evolution of the population of speeders can be determined from the current state and future controls of the system, the time at which such a crackdown would be necessary can be determined in advance as well.

5 Conclusion

We have shown how a principal can exploit myopic social learning to his advantage for a wide class of games where the interests of the population and the principal are directly opposed. In addition to the class of games we presented, the application domain of the methods we employed in this paper is vast. The basic idea is to indirectly influence decisions in the population through manipulating the payoffs associated with certain actions. Naturally, since the modified payoffs are not part of the initial system, such a manipulation comes at a personal and/or a social cost. In our example the principal had to expend an initial cost by either over-auditing or by letting the guilty population go unpunished. At the same time, there is a social cost to increased auditing that comes from the disutility honest agents obtain from being audited (the case where $v_2 > v_3$ in Figure 1). This initial phase however, is justified by later gains: since the population's reaction time to changes in the system is not instantaneous, the principal can revert back to the original game while the population *still plays as if they are in the modified game*. During this time, the principal enjoys a period of improved system performance. Generally, the solution either takes the form of a policy like the one we saw in this paper, where an initial period of extreme (in)activity is followed by a steady state, or it can be more cyclical in nature, with a cycle consisting of a phase that creates, via population learning, a certain impression about the environment followed by a phase where that impression is exploited. One obvious application is advertising. In this scenario, periods of heavy (and costly) advertising are followed by periods of relatively little advertising activity. During these latter periods, the effects from the initial advertising campaign continues to reverberate through the population, essentially providing free advertising until the effect dies down, at which time the advertiser starts the cycle again. A very different example is traffic regulation through periodic closures of specific roads. Such closures force drivers to change their driving habits.

Later, when these roads are re-opened, drivers take a while to adjust back to the initial equilibrium, as can be seen in [6]. Depending on the system's parameters, this lag in adjustment can provide the population with an average decrease in travel latency[3]. Applying the same approach of exploiting behavioral trends to other behavioral models would be an interesting next step in this line of research, with an eventual goal of cataloging the benefits that a principal or a society can obtain (or lose) as the level of sophistication of the population increases.

References

1. Acemoglu, D., Bimpikis, K., Ozdaglar, A.: Communication Dynamics in Endogenous Social Networks. Working Paper (2010)
2. Acemoglu, D., Dahleh, M., Lobel, I., Ozdaglar, A.E.: Bayesian learning in social networks. NBER Working Paper (2008)
3. Borgers, T., Sarin, R.: Learning through reinforcement and replicator dynamics. Journal of Economic Theory 77(1), 1–14 (1997)
4. Crawford, V.P., Kugler, T., Neeman, Z., Pauzner, A.: Behaviorally Optimal Auction Design: Examples and Observations. Journal of the European Economic Association 7(2-3), 377–387 (2009)
5. Eeckhout, J., Persico, N., Todd, P.: A Theory of Optimal Random Crackdowns. American Economic Review (2010)
6. Fischer, S., Vöcking, B.: On the evolution of selfish routing. In: Albers, S., Radzik, T. (eds.) ESA 2004. LNCS, vol. 3221, pp. 323–334. Springer, Heidelberg (2004)
7. Fudenberg, D., Levine, D.K.: The theory of learning in games. The MIT Press, Cambridge (1998)
8. Fudenberg, D., Maskin, E.: The folk theorem in repeated games with discounting or with incomplete information. Econometrica: Journal of the Econometric Society 54(3), 533–554 (1986)
9. Kamenica, E., Gentzkow, M.: Bayesian persuasion. NBER Working Paper (2009)
10. Myerson, R.B.: Optimal auction design. Mathematics of operations research 6(1), 58 (1981)
11. Nisan, N., Ronen, A.: Algorithmic mechanism design. Games and Economic Behavior 35(1-2), 166–196 (2001)

[3] For example, in Pigou's network with latencies 1 on the top link and x on the bottom link, a traffic authority can change the latency on the bottom link by slightly *increasing* it above 1 (say, by closing down a few lanes), so that the population migrates upwards towards the now-faster link. The latency is then restored back to its original value x, and the population starts migrating downwards. During that second phase, the traffic is temporarily more balanced than the Nash equilibrium, where everyone uses the bottom link all the time. The process is then repeated. The traffic authority chooses the exact latencies for the bottom link as well as the duration with which they remain in effect in order to obtain a net-gain in the average latency in the system.

Appendix

A Proof of Lemma 2

Proof. Denoting the function inside the integral in (9) by $L(t, x, \dot{x})$, the Euler-Lagrange equation gives another necessary condition that the optimal $x^*(t)$, if it exists, satisfies. Writing down the equation, we have

$$
\begin{aligned}
0 &= \frac{\partial L}{\partial x} - \frac{\partial}{\partial t}\frac{\partial L}{\partial \dot{x}} \\
&= e^{-rt}\left(c_3 + \frac{1}{p+q}(c_2 + (c_1 - c_2 - c_3)x(t))\left(\frac{\dot{x}(t)}{(1-x(t))x(t)^2} - \frac{\dot{x}(t)}{(1-x(t))^2 x(t)}\right)\right) \\
&+ e^{-rt}\left(\frac{(c_1 - c_2 - c_3)\left(p - \frac{\dot{x}(t)}{(1-x(t))x(t)}\right)}{p+q}\right) \\
&- e^{-rt}\frac{\left(r(-1+x(t))x(t)(-c_2 + (-c_1 + c_2 + c_3)x(t)) + \left(c_2 - 2c_2x(t) + (-c_1 + c_2 + c_3)x(t)^2\right)\dot{x}(t)\right)}{((p+q)(-1+x(t))^2 x(t)^2)}
\end{aligned}
$$

After some algebra and simplifying the above, we get

$$
\frac{e^{-rt}\left((c_2r - (c_1 + c_2)(p - r) + c_3(q + r))x(t) + ((c_1 - c_2)p + c_3q)x(t)^2\right)}{(p+q)(x(t) - 1)x(t)} = 0
$$

which is a quadratic function in $x(t)$. Solving that equation and enforcing the constraint that $0 \leq x(t) \leq 1$, we obtain the solution

$$
x^*(t) = \frac{(c_2 - c_1)p - c_3q + (c_1 - c_2 - c_3)r + \sqrt{4c_2((c_2 - c_1)p - c_3q)r + ((c_1 - c_2)p + c_3q + (c_1r - c_2 - c_3)r)^2}}{2((c_2 - c_1)p - c_3q)}
$$

which is time-independent and only depends on the parameters of the problem.

The difference between $x^*(t)$ and x_{Nash} depends on the parameters. For example, if we set all the parameters to 1 and compare the resulting steady state optimum with the Nash equilibrium in (1). We get

$$
x^*(t) = \frac{c_2}{c_3 + \sqrt{c_3^2 + c_2^2 - c_2(c_1 + c_3)}}
$$

which is always less than x_{Nash}.