DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

EXPLOITING MYOPIC LEARNING

Mohamed Mostagir

SOCIAL SCIENCE WORKING PAPER 1341

December 2010

Exploiting Myopic Learning^{*}

Mohamed Mostagir[†]

California Institute of Technology

November 23, 2010

Job Market Paper

Abstract:

I develop a framework in which 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 this framework 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. To illustrate the approach, I give a wide class of games for which the principal can always obtain strictly better outcomes than the corresponding Nash solution and show how such outcomes can be implemented. The framework is general enough to accommodate many scenarios, and powerful enough to generate predictions that agree with empirically-observed behavior.

JEL classification: *C72, C73, D03*

Keywords: Social Learning, Repeated Games, Bounded Rationality

^{*}I thank John Ledyard, R. Preston McAfee, Thomas Palfrey, and Jean-Laurent Rosenthal for their comments and suggestions. I also enjoyed many useful discussions with Dustin Beckett and Théophane Weber.

[†]Email: mosta@caltech.edu. Mailing Address: California Institute of Technology 1200 E. California Blvd, MC 228-77 Pasadena, CA 91125-0077.

1 Introduction

Repeated interactions between a principal and a population of agents are at the core of many fundamental models in economics, business, and politics. Most of these models consider the interaction between the principal and the agents in isolation, without accounting for the interactions amongst the agents themselves and how these interactions shape their decisions through social learning. At the same time, social learning research has witnessed a large boom, prompted in large part by the mounting evidence of its importance to business success, forming political opinions, and the spread of information and trends. The overwhelming majority of theoretical results in this area assume a population that learns in accordance with Bayes' rule. While these results are interesting in their own right and provide a useful benchmark, they disregard the voluminous amount of experimental evidence that suggests that people do not in fact seem to act in a Bayesian fashion.¹ This highlights the need for a) developing *behavioral* learning models that have the power to predict actual observed behavior and b) understanding how such models can be manipulated by a principal to maximize his gains. In this paper, I address both of these points in the context of a simple behavioral learning model.

The learning model I employ is that of replicator dynamics (Borgers and Sarin (1997)). 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 of how agents arrive at a Nash equilibrium. Under this model, a large 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 switches 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.² What makes replicator dynamics particularly appealing is that it is a simple

¹For example, see Tversky and Kahneman (1974), Tversky and Kahneman (1981), Camerer (1987), Griffin and Tversky (1992), and the surveys in Rabin (1998) and Camerer (1995).

²See, for example, Bomze (1986), Fischer and Vocking (2004), Fischer, Racke, and Vocking (2006), and the survey

form of learning dynamic that nicely straddles the line between behavioral and rational models. On one hand, agents update 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. Another 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. This underlies the fact that agents do not instantaneously react to their environment, and that switching to a new strategy is not always costless.

The central idea developed in this paper is that a principal can exert an important indirect influence on agents' decisions by exploiting their learning dynamics. I focus on games where the principal and the population's interests are diametrically opposed, though as I discuss later, the methods readily extend to a variety of other settings. I will give a formal definition of the class of games I consider in Section 2.1, but an informal description follows. There is a population where each member makes a choice from two pure actions. For simplicity, one 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 misreport their income or not, break the speed limit, accept a bribe or 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 the agent and run the risk of incurring a higher cost if the agent is cheating.³ Agents are interested in maximizing their payoffs, while the principal tries to minimize the costs from auditing and cheating. The game is repeated indefinitely. The principal's move in each round consists of choosing a fraction of the population to audit. As I will show, under 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. The question is whether the principal can improve on the

in Fudenberg and Levine (1998).

³One can think of non-policing scenarios that have a similar structure. For example, the agent can be a consumer faced with a choice of buying a product or not and the principal is a firm that decides whether to produce a low quality product or a high quality product at an extra cost. A firm prefers that consumers buy the low quality product so that it saves on production costs, and the consumer derives more value from buying a high quality product.

Nash outcome if the population learns according to replicator dynamics.

The primary contribution of this paper is twofold. On the conceptual front, I argue that imperfect decision making in a population –in its various formats– can in some cases be considered a resource that most system planners fail to utilize. The second contribution is methodological, where this abstract idea is implemented in the context of naive social learning. The main results of the paper can be summarized as follows:

1. If the principal himself is myopic, so that he reacts to the actions of the population without taking the future into account, then the interactions between the principal and the population leads to outcomes with a cyclical nature. As I discuss, such cycles are widely observed in the real world. This suggests that the learning model I consider captures essential elements of how people actually behave, and that the approach advanced in this paper not only provides a prescription for optimizing systems with a social learning component, but is also able to make positive predictions about how some existing systems actually operate.
2. By understanding the dynamics of the population and taking the future into account, the principal can obtain a higher payoff than that of the Nash equilibrium while doing *strictly less* auditing than what the Nash solution requires. A remarkable outcome, and one that achieves both goals set forth in the opening paragraph.

This paper is related to several strands of literature on (behavioral) mechanism design and social learning. Whether requiring an agent to update its information in a Bayesian fashion or to have perfect look-ahead and recall, standard economic theory endows the traditional rational agent with a set of abilities seldom found in human decision makers, and all the classic mechanism design results have been derived under that framework. There is however a recent stream of literature that studies agents under more realistic assumptions in order to match theoretical results with observed behavior. For example, Crawford and Iriberri (2007) argue that bidders who behave in accordance with the empirically-plausible level- k models (Stahl et al. (1994) and Stahl and Wilson (1995)) can explain overbidding and the winner’s curse in auctions. Crawford, Kugler, Neeman, and Pauzner (2009) give examples where it is possible under that model to obtain more revenue

than what is feasible under full rationality (Myerson (1981)). Earlier this decade, Nisan and Ronen (2001) launched the field of algorithmic mechanism design, an area that continues to thrive on questions of how the computational limits of decision makers affect their incentives as well as the outcomes obtained under the traditional agent models. In the same spirit as these works, the agents I consider here are not fully rational as their behavior is one of simple imitation. In addition, they base their decisions only on their most recent experience, paying no attention to their past history playing the game.

This paper also contributes to the recent work on social learning. The literature in this area has focused on when social learning 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. By letting the size of the population go to infinity, Acemoglu, Dahleh, Lobel, and Ozdaglar (2008) and Acemoglu, Bimpikis, and Ozdaglar (2009) derive limit results on the conditions under which a society can uncover the true state of the world. Because the society is assumed to be Bayesian, these models suffer from the criticisms laid out in the opening paragraph, namely that human agents seem to be unable to perform complicated belief updating procedures. Furthermore, unlike the model I study here, these models do not make predictions that can be corroborated by empirical observations. Understanding how a Bayesian society can be manipulated by a principal is a topic that has not yet been tackled in the social learning literature, though Kamenica and Gentzkow (2009) show how one can manipulate Bayesian agents, albeit outside of a social learning setting. The critical departure in this paper is the focus on both *behavioral* social learning and how it can be taken advantage of by a principal.

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 (Fudenberg and Maskin (1986)). Indeed, as alluded to earlier, for the class of games I 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.

The rest of the paper is organized as follows. Section 2 presents the class of games that I will focus on for the rest of the paper. Section 3 discusses the case of a myopic population *and* a myopic principal. Section 4 derives the optimal policies when the principal is forward-looking and Section 5 discusses how these policies improve over the rational population case. Section 6 provides empirical evidence that supports the predictions of the model. Section 7 gives other applications for the methods used in the paper. Section 8 quantifies how the degree of sophistication of the population impacts the principal, and Section 9 concludes the paper.

2 Model

I start by applying the ideas in the previous section to a class of 2×2 games that a large population repeatedly plays against a principal. In each round of the game an agent has one of two choices, a 'safe' choice with a low payoff, 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 taxation scenario, the costly action would be to audit an agent, and the 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 in this case expends auditing resources with no useful returns and – depending on how one sets up the model – can also incur a social cost in the form of the disutility or inconvenience that honest agents suffer because of auditing.

2.1 The Cheat-Audit Game

The example above is part of a large class of games that I call Cheat-Audit games. The payoffs of these game are as shown in Figure 1, with the principal being the column player. Each agent

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

Figure 1: The Cheat-Audit Game

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, I will assume that $v_1 = v_2$, so that an agent is indifferent to auditing as long as he is honest. This assumption has no impact on any of the structural results I obtain. 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 without catching a cheating agent (outcome (H, A)) is more costly than auditing a cheating 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 . I 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. That the principal's action consists of choosing a fraction to audit implicitly assumes *anonymity* of the agents in the population. In particular, the principal reacts to the *distribution* of play produced by the population, not the action of each

individual.⁴ I formalize this in the following assumption.

Assumption 2.1. (*Anonymity*) *All members of the population in the Cheat-Audit game look the same to the principal.*

The diametric opposition of the principal's and agents' interests implies that the game has no pure strategy equilibria, as indeed can be checked from Figure 1 and the relationship between the various payoffs. In fact, similar to a game of matching pennies, the single stage game possesses only a unique equilibrium in mixed strategies. Let the equilibrium audit rate and the fraction of C players in the fully rational setting be given by α_N and x_N , respectively. With the assumption that $v_1 = v_2$, it is straightforward to verify that

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

As mentioned, I consider this game in an infinitely-repeated setting where each moment in time the game in Figure 1 is played. Discrete time and how it affects the results I obtain is discussed in Section 6. I will let the state of the system at time t be the fraction of the population taking action C at that time, and will denote this fraction by $x(t)$. The principal's choice of audit rate at time t is denoted by $\alpha(t)$. The large population assumption together with anonymity immediately imply the following result.

Theorem 2.2. *The infinitely repeated Cheat-Audit game has a unique equilibrium in mixed strategies. This equilibrium is the same as that of the stage game.*

Proof. See Appendix. □

The reason why Theorem 2.2 is true is that, because each agent is a negligible part of the continuum, any individual action has no effect on the distribution of play and thus no bearing on the future treatment of that agent.

⁴One can think of scenarios where the identity of the player can be useful in the punishment phase, but not the detection phase. For example, police does not observe the license plate of a vehicle and then decide whether or not to apply a radar gun. The fact that a speeder is a repeat offender only comes into play after he has been caught and does not affect his probability of being detected.

Given a state $x(t)$, audit rate $\alpha(t)$, and denoting the payoff to the principal at time t by $g(t)$, the cost to the principal at time t is given by

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

where the terms in the first equation 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 strategies and payoffs. Since agents only switch strategies if they meet someone who is playing a different strategy from their own, switching can only happen if the share of the different strategies in the population is initially positive, otherwise everyone will continue to play the same strategy forever. The following assumption is therefore essential.

Assumption 2.3. *At the beginning of the horizon each strategy is played by a positive share in the population, i.e. $0 < x(0) < 1$.*

Under this 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 $k > 0$. One can think of k as a 'speed of transmission': the willingness of an agent to change their strategy when they meet someone with a better experience. Without loss of generality, I will assume that an agent who obtains payoff u switches to the strategy of an agent who obtained 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 I obtain. Any scheme where the switching probabilities are proportional to the payoff differences, so that the share of strategies that perform better grows in the population, essentially leads to the same results. I will make the derivations less cumbersome and more general by assuming that switching in the first scenario happens with probability p and in the second scenario with probability q , and 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 . I can then write the dynamics of the system as a function of $x(t)$ and $\alpha(t)$

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

3 Myopic Principal

Before discussing the optimal policy for the principal, I consider the following question: what happens if the principal himself is myopic? Although there is strong reason to believe that the principal is more sophisticated than the population, there are many scenarios that encourage a short-sighted principal. A politician can pander to an electorate in the hopes of obtaining an immediate reward, or a corporate manager can make decisions with the goal of improving short-term gains as a response to pressure from investors. I will analyze such situations in this section by assuming that the principal himself learns in a myopic fashion and that he does this by adjusting his strategy after each round of the game in a similar manner to the population. This necessitates an assumption similar to Assumption 2.3.

Assumption 3.1. *At the beginning of the horizon the principal assigns positive weights to each of his strategies, i.e. $0 < \alpha(0) < 1$.*

Like the previous section, the cost of action α is $c_1\alpha x + c_2\alpha(1 - x)$ while the cost to $(1 - \alpha)$ is equal to the cost of those cheating agents who went away undetected, and is equal to $c_3(1 - \alpha)x$. After each round, the principal observes the costs from both actions H and I and adjusts the proportion by which he plays them in the next round according to how well they did in the current round. Of course, the principal has no way of knowing whether the members of the population who were not audited were cheating or not. This is easily overcome by the large population assumption, since the fraction of the population that the principal audits identifies the fraction of cheaters in the population with probability one, and this fraction can then be used to estimate the costs incurred from not auditing.

It turns out that when the principal is also myopic, the system oscillates: periods of high cheating activity induce periods of intense monitoring activity by the principal. This high-intensity auditing in turn drives the population to periods with little or no cheating activity and consequently, leads the principal to perform less auditing. This pattern continues indefinitely in a cyclic fashion. The unique Nash equilibrium of (1) is an unstable equilibrium, or center, of this dynamical system. This means that even if the system starts at equilibrium, any small perturbation will send it into oscillation. The following result gives a more precise description of the nature of interaction between the principal and the population under this setup.

Theorem 3.2. *The fluctuations of the population of C players and the audit rate of the principal are periodic. The period depends on the values of the problem as well as the initial conditions. The unique mixed equilibrium of (1) is a center of the dynamical system induced by the repeated play of the Cheat-Audit game.*

Proof. See Appendix. □

Informally, the equations with which the fraction of C players and the audit rate evolve describe a dynamical system with a unique non-hyperbolic equilibrium. This equilibrium corresponds to the Nash equilibrium in (1), and – because the system has only two eigenvalues on the imaginary axis–

is also a center of the system. This means that small perturbations push the system away from equilibrium. Being a center also implies that the path of any solution to the dynamical system is a closed orbit around the center. Thus, the system revisits each point in its evolution periodically.

Figure 2 displays a phase portrait of the system, with the fraction of C players on the x -axis and the audit rate on the y -axis. The closed orbit represents a solution that satisfies (18) and (19). As Theorem 3.2 implies, the principal's audit peaks trail the peaks of the fraction of cheaters in the population, leading to a cyclical nature in both the audit activity as well as the size of the population of C players. Suppose, as in the figure, that the system starts from a point in the interior of the unit square with low cheating and auditing activities, then the lack of policing encourages the population to cheat, since agents learn that cheating is the action that provides a higher payoff. As the number of cheaters increases, the principal starts to ramp up the auditing activity, leading to extreme auditing of the population that eventually drives the majority to play H again, and the cycle repeats. As I discuss in Section 6, this cyclical nature can be observed in various real-world phenomena that correspond to the Cheat-Audit game.

3.1 Average Cheating and Audit Rates

It is natural to ask how the scenario analyzed above differs from the rational case. It turns out that if the game is played long enough, then the players' actions, averaged over time, are equal to the corresponding values in the fully-rational setting. This accentuates the early discussion about replicator dynamics: even though they are following very simple rules, both the principal and the agents are able to approximate the behavior of their rational counterparts. In particular, the fraction of C players and the principal's audit rate over any period are the same as those obtained in the mixed equilibrium solution given by (1). The following result formalizes this fact.

Theorem 3.3. *The average audit rate of the principal and the average fraction of cheaters over any period are the same as the corresponding Nash equilibrium values.*

Proof. See Appendix. □

Having shown that the outcome of the game between the myopic principal and the population is

Figure 2: Phase Portrait of Cheating and Auditing Activity

close to the fully-rational outcome, I proceed to show how a forward-looking principal can improve on the results of the fully-rational setting.

4 Forward-Looking Principal

A forward looking principal differs from the myopic principal of the last section in that he takes the future into account, so that instead of reacting to the latest round of the game, he optimizes over the (infinite) horizon of the problem. In the following I define the principal's objective and derive the optimal policy that achieves it.

4.1 Objective

The principal's problem is the following. Given the different values in Figure 1 and the learning dynamics, the principal is interested in minimizing his long-run discounted cost. This long-run cost is the discounted sum of all costs accrued from playing the game over time. Recall that the payoff

at time t is given by (2) and the equation of motion of the population by (3). The principal's problem can then be written as

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

where $r > 0$ 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 $\alpha(t)$.

4.2 Optimal Policy

4.2.1 Single Round

Before trying to find the optimal solution to (4), I will first develop an intuition by considering the solution when the game is played only once. The stage game cost described by (2) can be factored and rewritten as

$$g(x, \alpha) = c_3x + \alpha(c_2 + (c_1 - c_2 - c_3)x)$$

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 above 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 :

$$g^*(x) = \begin{cases} c_3x, & x < \frac{c_2}{c_2+c_3-c_1}; \\ c_2 + (c_1 - c_2)x, & x \geq \frac{c_2}{c_2+c_3-c_1}. \end{cases} \quad (6)$$

As I will show later, 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.

4.2.2 General Policy

I will derive the optimal policy for (4) by formulating the Hamiltonian function for the system and using the Euler-Lagrange equation. I 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 one 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

$$\begin{aligned} H(x, \alpha, \lambda) &= g(x, \alpha) + \lambda f(x, \alpha) \\ &= c_3x + \alpha(c_2 + (c_1 - c_2 - c_3)x) + \lambda x(1 - x)(q - \alpha(q + p)) \\ &= c_3x + \lambda qx(1 - x) + \alpha(c_2 + (c_1 - c_2 - c_3)x - \lambda(p + q)x(1 - x)) \end{aligned} \quad (7)$$

where λ is a co-state variable that represents 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 essentially reduces the general problem to a single period one. The following lemma utilizes the Hamiltonian to provide necessary, but not sufficient, conditions on the optimal control trajectory.

Lemma 4.1. *The optimal control for Problem (4) is a bang-bang solution, where*

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

Proof. See Appendix. □

Lemma 4.1 implies that the optimal control function, $\alpha^*(t)$, takes values on its boundaries. 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.⁵ This provides some information about the structure of the optimal policy, but not enough to completely characterize it. To do this, I will reformulate (4) as a calculus of variations problem. From (3), I can express $\alpha(t)$ as

$$\alpha(t) = \frac{1}{p+q} \left(p - \frac{\dot{x}(t)}{x(t)(1-x(t))} \right) \quad (9)$$

Substituting this into the objective, the problem becomes

$$\begin{aligned} & \min_{x(t)} \int_0^\infty e^{-rt} g \left(x(t), \frac{1}{p+q} \left(p - \frac{\dot{x}(t)}{x(t)(1-x(t))} \right) \right) dt \\ &= \min_{x(t)} \int_0^\infty e^{-rt} \left(c_3 x(t) + \frac{(c_2 + (c_1 - c_2 - c_3)x(t)) \left(p - \frac{\dot{x}(t)}{(1-x(t))x(t)} \right)}{p+q} \right) dt \end{aligned} \quad (10)$$

The solution to (10) provides another necessary condition on the optimal state trajectory. Specifically, the following lemma shows that there is a constant for which the integral in (10) is stationary, i.e. the function that minimizes (10) is time-independent.

⁵As I show in the proof of Theorem 4.3, the third case happens only for precisely a single pair $(x^*(t), \alpha^*(t))$ in the optimal solution.

Lemma 4.2. *Let $x^*(t)$ be the minimizer to (10), then $x^*(t) = C$, where C is a constant that depends on the parameters of the problem and is equal to*

$$\frac{(c_2 - c_1)p - c_3q + (c_1 - c_2 - c_3)r + \sqrt{4c_2((c_2 - c_1)p - c_3q)r + ((c_1 - c_2)p + c_3q + (c_1r - c_2 - c_3)r)^2}}{2((c_2 - c_1)p - c_3q)} \quad (11)$$

Proof. See Appendix. □

The necessary conditions I have obtained so far are enough to fully characterize the optimal policy.

Theorem 4.3. *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.*

Theorem 4.3 indicates that, depending on a threshold value, the optimal solution either audits indiscriminately or does nothing. If the system hits the value $\bar{x}$, then the system audits at a constant rate. The structure of the optimal policy then is quite different from the myopic principal case, where the principal's strategy oscillates continuously. As I discuss in Section 6, the optimal policy can oscillate too, as a result of considering the model in discrete time.

5 Comparison With The Nash Equilibrium

How does the solution for the class of games considered here fare under the forward-looking principal in comparison to the fully rational Nash equilibrium outcome? I have already discussed in Section 2.1 and in Theorem 2.2 that the (fully rational) repeated game possesses a unique equilibrium in mixed strategies, given by (1). As I have shown in Theorem 3.2, this equilibrium is also a *center* of the repeated behavioral game. This means that, under the replicator assumption, there exists 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, i.e. the principal can implement the Nash outcome in the behavioral setting, if he so desires. However, the optimal solution that I

obtained in Section 4 is not the Nash equilibrium, indicating that the Nash solution is dominated by the policy in Theorem 4.3. Furthermore, as I show below, as soon as the game reaches steady state, the optimal policy involves *less* auditing than the Nash solution. Because of this, the Nash solution never coincides with the policy in Theorem 4.3, 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.

Beyond the audit rate, It is also instructive to look at how the fraction of C players compares under the behavioral and the rational settings. The following theorem summarizes the results that a principal can obtain when facing a behavioral population. The principal is able to both perform less auditing and, if he cares about the future enough, keep the fraction of C players close to zero.

Theorem 5.1. *The steady state audit rate in the behavioral setting is strictly less than the Nash audit rate. Let $r > 0$ be a discount factor, then $\lim_{r \rightarrow 0} \bar{x} = 0$, i.e. as the principal cares more about the future, the fraction of C players is driven close to zero. This contrasts with the Nash fraction of C players x_N , which is insensitive to the effect of discounting.*

Proof. See Appendix. □

This result highlights the stark difference between the behavioral and rational settings. Discounting has no bearing on the outcome in the rational case since, as Theorem 2.2 shows, the principal cannot influence the future actions of the population. Furthermore, from Theorem 3.3, the outcomes in the behavioral setting are close to the Nash solution when the principal responds myopically. By taking the future into account however, the principal is able to obtain outcomes that were not possible under these other scenarios. As I discuss in the next section, the results in Theorems 3.2 and 4.3 are widely observed in practice.

6 Examples

Eeckhout, Persico, and Todd (2010) define crackdowns as intermittent periods of high-intensity monitoring. Crackdown cycles occur when these periods are interweaved with periods of lax enforcement. There is a wealth of examples of this phenomenon. Di Tella and Schargrodsky (2003)

study crackdowns on corruption in hospitals in Buenos Aires, Lui (1986) describes crackdowns on corruption in China, the Recording Industry Association of America (RIAA) utilizes crackdowns to combat illegal file sharing, and police in Belgium intermittently cracks down on speeders. I discuss some of these examples and show how they relate to the results of the previous sections.

The paramount example of crackdown cycles is how the Chinese government controls corruption and dissidence. Lui (1986) describes three major crackdowns in China over the period from 1950 to 1982. The first campaign, known as the *san fan*, started in 1950 and lasted for two years, ending in June of 1952. The campaign was characterized by a highly intensive effort that managed to reduce crime from 500,000 cases in 1950 to an average of 290,000 cases over the following 15 years. The *san fan* was not just characterized by severe punishments, but also by extremely high auditing activity, and during the crackdown period crime steadily declined to very low rates. As the cycle in Figure 2 predicts, the post-crackdown period was characterized by low crime rates *and* low monitoring activity, and has nowadays come to be known as “the golden age of honesty” in China⁶, where as Lui puts it, “the Chinese government did not spend any significant amount of resources on auditing”. Eventually though, corruption started to increase again, and the government cracked down on both corruption and dissidence in the middle of the 1960s. The pattern was then repeated as the decrease in monitoring after the second crackdown led to a rise in corruption levels, which by 1979 were getting out of control. This led to a third crackdown that started in 1982 and lasted for more than three years.

A more recent example is how the RIAA and the Motion Picture Association of America (MPAA) fight online piracy and illegal file sharing. Figure 3 shows copyright infringement lawsuits in the United States over the period 1993-2009.⁷ The beginning of the millennium witnessed a huge increase in the number of file sharers, where platforms like Napster had a record 26 Million users at one point. The percentage of internet users who were also illegal file sharers continued to grow, hitting a high of 29% of all US internet users.⁸ As one can see in the figure, the RIAA responded with a severe crackdown that started around 2004 and lasted for five years. During the crackdown,

⁶Lui (1986)

⁷Source: Administrative Office Of The Courts.

⁸Source: PEW Internet and American Life Project Data Memo.

the amount of infringement lawsuits tripled. Most of these lawsuits targeted anonymous, 'John Doe' defendants. The crackdowns resulted in a drop in the percentage of file sharers from 29% to 14%, with the number stabilizing somewhere around 18%.⁹ In 2008, the RIAA announced that it has stopped its mass-lawsuit practice but that it will continue to sue users at a lesser rate. Although in 2010 it is early to tell, this pattern bears a striking resemblance to the policy in Theorem 4.3, where a severe crackdown brings the fraction of offenders down to a certain level, after which auditing continues at a lower rate. Of course, there are many factors that go into a campaign like the one launched by the RIAA, including publicity of, and backlash against, the lawsuits, but the overall agreement of the pattern with the results obtained in this paper suggests that the core driving factors are captured by the model.

The analysis in this paper focused on a continuous time framework. In reality, many of the games that fit the model take place in discrete time, or require resources that are infeasible to implement forever as the optimal solution requires. 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 $x(t)$ down to its optimal value. Because of the discreteness, the crackdowns always bring the value of $x(t)$ below $\bar{x}$, hence leading to periods of low activity on the principal's part. The whole cycle is then repeated as $x(t)$ increases above the threshold again. Eeckhout, Persico, and Todd (2010) empirically observe crackdowns by the police on speeders in Belgium. They give a static model with a rational agent population and show that under some assumptions a crackdown can be part of the optimal strategy of the police. While they note the periodicity of the crackdowns, their static model is unable to provide an explanation for this phenomenon. Additionally, through anecdotal evidence and conversations with the police, they mention that crackdowns are planned as early as a month in advance. Both of these observations are explained by the model in this paper. 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

⁹However, the *amount* of copyrighted material shared online continues to grow.

Figure 3: US Copyright Infringement Cases 1993-2009

well.

7 Other Applications

7.1 Equilibrium Selection and Technology Adoption

The framework I use in this paper can also be used as a device for equilibrium selection. The fact that a game may possess multiple equilibria makes it more difficult to design mechanisms that select for a particular equilibrium with certain desired outcomes. Balcan, Blum, and Mansour (2009) consider the problem of moving a population from one equilibrium to another one with more socially desirable properties. Their framework uses public advertisement as a means to influence decisions in the rational agent population, and they analyze the effectiveness of this method even when only a small fraction of the population follows the advertisement. In many cases, the proposed method fails to move the population between equilibria. For coordination games like the one in Figure 4, I show how a principal can steer the population towards an equilibrium that is worse for them but is beneficial for the principal.

As an example, assume a firm wants to replace an old technology with a new one (for example,

	L	R
T	a_1, b_1	0,0
B	0,0	a_2, b_2

Figure 4: Coordination Game

a new version of a mobile device operating system). The firm prefers that users switch to the new technology since they would have to update their devices as well as have the capability to run improved and more expensive services, resulting in more revenue for the firm. The population, however, has no strong inclination to do the switch as long as the existing platform is supported. Because the firm has limited resources, it can only split its support over the existing and the new versions of the technology, generating ill-will amongst consumers who are not receiving proper support and perhaps risking that these consumers abandon the product or service altogether.

The situation above is depicted in the coordination game in Figure 4, with the principal being the column player. Assume $a_1 < a_2$ and $b_1 > b_2$, so that the principal's preferred outcome is (T, L) , while the population prefers outcome (B, R) . Similar to the setup in Section 2, I denote the fraction with which the principal plays action L at time t by $\alpha(t)$, while the fraction of the population playing L is denoted by $x(t)$. One can interpret α as the fraction of the firm's resources that it devotes to sustaining technology L and $1-x(t)$ as the fraction sticking with the old technology. Assumption 2.3, the share of the population playing either strategy at the beginning of the horizon is positive. This corresponds to the fact that the new technology has early adopters (). The question then is whether the principal can facilitate the migration of the population towards an equilibrium that is more desirable for him, which in Figure 4 corresponds to equilibrium (T, L) .

As before, I will first consider the outcome of this interaction when the principal adopts a myopic approach. For scenarios like the one discussed above, it is reasonable to expect that the

system starts somewhere close to the (B, R) equilibrium, where most of the population has still not adopted the new technology and the firm still offers extensive support for the old technology. Under this setting, a myopic principal cannot move the population near equilibrium (T, L) . In fact, the following result shows that a myopic principal gets stuck at the (B, R) equilibrium forever.

Theorem 7.1. *If the system starts in an interior state where $x(0) > \frac{b_1}{b_1+b_2}$ and $\alpha(0) > \frac{a_2}{a_1+a_2}$, then myopic learning on both the population and the principal sides converges to the (B, R) equilibrium of the game in Figure 4.*

Proof. See Appendix. □

Thus a myopic principal's outcome crucially depends on the starting point of the system. This is a direct result of the reactionary nature of this type of principal: if the population deems the target equilibrium as undesirable, then the firm will not risk alienating its customers by trying to move them to that equilibrium.

Can a forward-looking principal do better? Using the same terminology from Section 2 to formulate the dynamic problem, the principal's payoff at time t is $g(x(t), \alpha(t)) = \alpha(t)(1 - x(t)) + (1 - \alpha(t))x(t)$. Like (4), the principal wants to solve

$$\begin{aligned} \min_{\alpha(t)} \quad & \int_0^\infty e^{-rt} (\alpha(t)(1 - x(t)) + (1 - \alpha(t))x(t)) dt \\ \text{s.t.} \quad & \dot{x}(t) = x(t)(1 - x(t))((1 - \alpha(t))a_2 - \alpha(t)a_1) \\ & 0 \leq \alpha(t) \leq 1 \end{aligned} \tag{12}$$

If this game is played only once, then the principal's optimal α is given by a cut-off strategy similar to the one period solution in (5), where

$$\alpha^* = \begin{cases} 0, & x < \frac{b_1}{b_1+b_2}; \\ 1, & x \geq \frac{b_1}{b_1+b_2}. \end{cases} \tag{13}$$

In the same way as in Lemma 4.1 and Theorem 4.3, Problem (12) has a unique optimal solution. The next result shows that this solution coincides with the equilibrium (T, L) when the discount

factor is high enough.

Theorem 7.2. *Under Assumption 2.3 and as $r \rightarrow 0$, a forward-looking principal has a strategy that converges to the (T, L) equilibrium of the coordination game in Figure 4.*

Proof. See Appendix. □

As I show in the proof of the theorem, the principal's optimal strategy essentially offers no support for the old product, leading the population to migrate towards using the new one.¹⁰ As the principal cares more about the future, he is willing to sustain some 'transitory costs', the costs that it incurs as a result of unsatisfied customers, in order to accumulate later gains. In this sense the optimal policy is similar to the policy in the Cheat-Audit game: crackdown on the population in the optimal strategy incurs costs from auditing H players, or not auditing anyone incurs costs from letting C players go undetected. These costs are, again, justified by the later gains to the principal.

8 Robustness

In contrast with classical economic theory, I have concerned myself with an agent population that is completely behavioral, and I compared my findings under different models of the principal to the predictions of the classical model. Under the behavioral mode, I have shown that the principal can implement more favorable outcomes than under the rational one. It is reasonable to expect that a more accurate model of the population lies in between these two extremes. How robust are the results to a population that consists of both rational and behavioral agents? As one would expect, the payoffs that a principal can sustain decrease as the population becomes more sophisticated. This section formalizes this fact in the context of Cheat-Audit games.

I assume that there are two types of agents in the population, a rational type, exemplified by the classical rational agent, and a behavioral type that acts in the same way as the agents I

¹⁰In reality, a firm is unlikely to stop support for an old product as soon as a newer version is released. However, it is not uncommon that the company substitutes lack of support with making the older product less functional. As an example, When Apple released the iPhone 4 and its accompanying OS, it effectively rendered the older 3G model slow and frustrating to use. Since newer applications required an upgrade to the new OS, a customer was faced with either decreasing the pool of applications that he can choose from, or upgrading to the new OS and experiencing the sluggish performance. Official Apple support was nonexistent for either scenario. This immediately decreased the utility of the old model.

have so far considered in this paper. Formally, let the fraction of rational agents be given by ρ , and let the state of the system at time t , $x(t)$ denote the number of C players *in the behavioral population only*. The population dynamics operate in a similar fashion as before, except that the rational population makes its choices in a strategic fashion as opposed to the more myopic approach of the behavioral types. The rational population is aware of its size and of the existence of the behavioral population, whereas the behavioral population is oblivious to that distinction. The principal knows the distribution of the population as well as the information that each population has about its environment. Under this setting, Theorem 2.2 still applies to the behavior of the rational population: the rational agents still optimize for the current period only. The reasoning is the same as in the proof of that theorem. A rational agent will best reply to the principal's action in this period because that agent's action has no consequences for the future, given the agent's size relative to the population and the anonymity assumption. With that in mind, I examine the best reply dynamics of the rational agents for the one-shot version of this game.

8.1 Single Round

The probability with which rational agents cheat depends on the principal's action α , which in turn depends on x and ρ . The following result summarizes how these agents behave in that setting.

Proposition 8.1. *Let ρ be the fraction of the rational population, x be the fraction of the behavioral population playing action C , and p_c be the probability with which a member of the rational population plays C . The optimal p_c , as a function of x is described by*

$$p_c^*(x) = \begin{cases} 0, & x > x_N; \\ 1, & x + \rho < x_N; \\ \frac{x_N - x}{\rho}, & x + \rho \geq x_N. \end{cases} \quad (14)$$

where x_N is as given in (1).

Proof. See Appendix. □

The reason the rational agents behave in this manner can be inferred from the principal's optimal

strategy in Section 4.2.1. When the population is purely behavioral, he uses a threshold rule to audit or not. If $x(t)$ is known to both the principal and the rational population, then the rational players know that if the total fraction of C players, given by the sum of the fraction of behavioral and rational players playing C is less than the threshold, then the principal's optimal action will be to not audit. Conversely, when the total is over the threshold the principal audits everyone. Similarly, the principal adjusts his action taking the reasoning of the rational population into account. This means that the only optimal strategy for the situation when the potential total number of C players is over the threshold is for both the principal and the rational players to play a strategy that makes the other indifferent. Since the principal is indifferent exactly at the threshold value, the rational population mixes with a probability that pushes the expected total number of C players to that value, and the principal mixes with a strategy that encourages the population to mix at that rate.

8.2 Multi-Round

Assume now that the members of the rational population, when meeting members of the behavioral population, do not lie about their experience. The fraction of the behavioral population then changes as before, except that behavioral agents meet other behavioral agents as well as rational agents. Because the rational population is able to figure out the principal's optimal strategy, their action at time t is a function of the principal's action, so that I can write the probability in Lemma 8.1 as $p_c^*(\alpha(t))$. Formally, the rate of change $\dot{x}(t)$ is given by

$$\dot{x}(t) = (1 - x(t) - \rho)(x(t) + p_c^*(\alpha(t))\rho)(1 - \alpha(t))q - \alpha(t)x(t)(1 - x(t) + (1 - p_c^*(\alpha(t)))\rho)p \quad (15)$$

where q and p are as before. The first term is the fraction of honest agents that switches to cheating when they meet cheating agents, whether rational or behavioral, who have not been audited. If $\alpha(t)$ is such that $p_c^*(\alpha(t)) = 0$ then the term reduces to the case we dealt with in Section 4.2.2. Similarly, the second term describes the fraction of those agents who were cheating and got caught and later meet honest agents, where again it does not matter whether the honest agents were behavioral or not.

The principal's problem then becomes

$$\begin{aligned}
\min_{\alpha(t)} \quad & \int_0^\infty e^{-rt} (c_1 - c_2 - c_3) \alpha(t) (x(t) + p_c^*(\alpha(t)) \rho) + c_2 \alpha(t) + c_3 (x(t) + p_c^*(\alpha(t)) \rho) dt \\
\text{s.t.} \quad & \dot{x}(t) = (1 - x(t) - \rho)(x(t) + p_c^*(\alpha(t)) \rho)(1 - \alpha(t))q - \alpha(t)x(t)(1 - x(t) + (1 - p_c^*(\alpha(t)))\rho)p
\end{aligned} \tag{16}$$

Writing the Hamiltonian for this problem,

$$\begin{aligned}
H(x, \alpha, \lambda) &= g(x, \alpha) + \lambda f(x, \alpha) \\
&= (c_1 - c_2 - c_3) \alpha (x + p_c^*(\alpha) \rho) + c_2 \alpha + c_3 (x + p_c^*(\alpha) \rho) \\
&\quad + \lambda \left((1 - x - \rho)(x + p_c^*(\alpha) \rho)(1 - \alpha)q - \alpha x (1 - x + (1 - p_c^*(\alpha)) \rho) p \right)
\end{aligned} \tag{17}$$

This is a more difficult problem than the one in (4). The difficulty comes from the fact that the Hamiltonian is no longer a linear function of α . In fact, the Hamiltonian function is not continuous anymore, since small changes in α near the stage Nash equilibrium value α_N can trigger p_c^* to take extreme values as in Lemma 8.1. Because of this, I will concern myself with more qualitative than quantitative issues when it comes to the optimal policy for this scenario. In particular, I show in Proposition 8.2 that if $\rho < x_N$, then the principal can audit strictly less than α_N while keeping the fraction of C players the same as in the Nash solution. Proposition 8.3 shows that the whole population does not have to be rational in order for the principal to audit with a rate that is at least α_N ; it suffices for the fraction of rational agents ρ to be more than x_N for that to happen. This implies that, as expected, the behavior of the principal is monotonic in ρ , with the auditing activity increasing as ρ increases.

Proposition 8.2. *Let $\alpha^*(t)$ denote the principal's optimal audit rate. If $\rho < x_N$, then $\alpha(t)^* < \alpha_N$.*

Proof. See Appendix. □

Proposition 8.3. *If $\rho > x_N$, then $\alpha(t)^* \geq \alpha_N$.*

Proof. See Appendix. □

9 Conclusion

This paper discusses a behavioral social learning model based on replicator dynamics. In this model, agents play a game repeatedly and switch between strategies based on which strategies are performing better in the population. I show that this model provides positive predictions of many observed phenomena in the real world, like the existence of police crackdowns and the cyclical nature of anti-corruption campaigns. In addition to the predictive power of the model, it provides a framework for a forward-looking principal to implement outcomes that are not possible under traditional rationality assumptions. The basic idea is that the principal can indirectly influence decisions in the population by manipulating the payoffs associated with certain actions over time. In the context of Cheat-Audit games, the principal is able to do better than the corresponding Nash equilibrium of these games and being able to do so with less auditing effort. In coordination games, I show how the principal can manipulate the population and influence them to migrate from one equilibrium towards another, more desirable one.

The application areas of the methods developed in this paper are vast. Advertising is one potential application where periods of heavy and costly advertising activity are followed by periods with relatively little advertising. During these latter periods, the effects from the initial advertising campaign continue to reverberate through the population, essentially providing free advertising until the effect dies down, at which time the advertiser starts the cycle again. A different example is traffic regulation through periodic closures of specific roads or periodic toll increases. Such changes force drivers to modify their driving habits. Later, when these roads are re-opened or tolls are reduced again, drivers take a while to adjust back to the initial equilibrium, as can be seen in Fischer and Vocking (2004). This lag in adjustment can be exploited to try and balance traffic over the available routes. On the other hand, there are games that are not prone to the framework presented in this paper. On the other hand, The Prisoner's Dilemma is one game where a principal does not have any strategy that would generate a higher payoff against a myopic population as compared to a rational population, since social learning will always lead the population to defect against the principal. It will thus be instructive to further understand the general features that

determine when one can exploit learning in a population.

Finally, the results in this paper show the promise of behavioral models as descriptive tools of reality and frameworks for optimization. Exploring other behavioral trends and understanding their explanatory powers and how they can be manipulated would be the natural next step in this line of research. There already exist established models of bounded-memory agents (for example, Young (1993)) that can be used as a foundation for work similar to the one in this paper. More recent models of behavioral qualities like thinking aversion (Ortoleva (2008)) can also be utilized as a starting point for designing behavioral mechanisms that exploit computational complexity in order to steer agents to choose alternatives that are easier to compute. Ultimately, while a unified theory of behavior seems unlikely or at least unattainable in the immediate future, the insights gained from studying various behavioral effects in isolation will undoubtedly contribute to a better understanding of their relative importance within such a theory and to the process of human decision making in general.

References

- ACEMOGLU, D., K. BIMPIKIS, AND A. OZDAGLAR (2009): “Communication Dynamics in Endogenous Social Networks,” .
- ACEMOGLU, D., M. DAHLEH, I. LOBEL, AND A. OZDAGLAR (2008): “Bayesian learning in social networks,” *NBER Working Paper*.
- BALCAN, M., A. BLUM, AND Y. MANSOUR (2009): “Improved equilibria via public service advertising,” in *Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 728–737. Society for Industrial and Applied Mathematics.
- BOMZE, I. (1986): “Non-cooperative two-person games in biology: a classification,” *International journal of game theory*, 15(1), 31–57.
- BORGERS, T., AND R. SARIN (1997): “Learning through reinforcement and replicator dynamics,” *Journal of Economic Theory*, 77(1), 1–14.
- CAMERER, C. (1987): “Do biases in probability judgment matter in markets? Experimental evidence,” *The American Economic Review*, 77(5), 981–997.
- CAMERER, C. (1995): “Individual decision making,” *The handbook of experimental economics*, 3, 587–704.
- CRAWFORD, V., AND N. IRIBERRI (2007): “Level-k Auctions: Can a Nonequilibrium Model of Strategic Thinking Explain the Winner’s Curse and Overbidding in Private-Value Auctions?,” *Econometrica*, 75(6), 1721–1770.
- CRAWFORD, V., T. KUGLER, Z. NEEMAN, AND A. PAUZNER (2009): “Behaviorally Optimal Auction Design: Examples and Observations,” *Journal of the European Economic Association*, 7(2-3), 377–387.
- DI TELLA, R., AND E. SCHARGRODSKY (2003): “The role of wages and auditing during a crack-down on corruption in the city of Buenos Aires,” *Journal of Law and Economics*, pp. 269–292.

- EECKHOUT, J., N. PERSICO, AND P. TODD (2010): “A Theory of Optimal Random Crackdowns,” *American Economic Review*.
- FISCHER, S., H. RACKE, AND B. VOCKING (2006): “Fast convergence to Wardrop equilibria by adaptive sampling methods,” in *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, p. 662. ACM.
- FISCHER, S., AND B. VOCKING (2004): “On the evolution of selfish routing,” *Algorithms-ESA 2004*, pp. 323–334.
- FUDENBERG, D., AND D. LEVINE (1998): *The theory of learning in games*. The MIT press.
- FUDENBERG, D., AND E. MASKIN (1986): “The folk theorem in repeated games with discounting or with incomplete information,” *Econometrica: Journal of the Econometric Society*, 54(3), 533–554.
- GRIFFIN, D., AND A. TVERSKY (1992): “The weighing of evidence and the determinants of confidence* 1,” *Cognitive Psychology*, 24(3), 411–435.
- KAMENICA, E., AND M. GENTZKOW (2009): “Bayesian persuasion,” *NBER Working Paper*.
- LUI, F. (1986): “A dynamic model of corruption deterrence,” *Journal of public Economics*, 31(2), 215–236.
- MYERSON, R. (1981): “Optimal auction design,” *Mathematics of operations research*, 6(1), 58.
- NISAN, N., AND A. RONEN (2001): “Algorithmic mechanism design,” *Games and Economic Behavior*, 35(1-2), 166–196.
- ORTOLEVA, P. (2008): “The price of flexibility: towards a theory of thinking aversion,” .
- PERKO, L. (2001): *Differential equations and dynamical systems*. Springer Verlag.
- RABIN, M. (1998): “Psychology and economics,” *Journal of economic literature*, 36(1), 11–46.

- STAHL, D., ET AL. (1994): “Experimental evidence on players’ models of other players,” *Journal of economic behavior & organization*, 25(3), 309–327.
- STAHL, D., AND P. WILSON (1995): “On Players Models of Other Players: Theory and Experimental Evidence,” *Games and Economic Behavior*, 10(1), 218–254.
- TVERSKY, A., AND D. KAHNEMAN (1974): “Judgment Under Uncertainty: Heuristics and Biases, Science,” *New Series*, 185(4157), 1124–1131.
- (1981): “The framing of decisions and the psychology of choice,” *Science*, 211(4481), 453.
- YOUNG, H. (1993): “The evolution of conventions,” *Econometrica: Journal of the Econometric Society*, 61(1), 57–84.

Appendix

Proof of Theorem 2.2

Proof. Assumption 2.1 implies that the principal can only respond to the distribution of play in the population. Because each agent is a negligible part of the continuum, any individual action has no effect on the distribution of play and thus no bearing on the future treatment of that agent, i.e. the continuation payoff for any agent is unaffected by its action in the current round. Hence it is optimal for an agent to play C if $E[C|\alpha(t)] > E[H|\alpha(t)]$, which is always the case when $\alpha(t) < \alpha_N$. The opposite is true when $\alpha(t) > \alpha_N$. Similarly, the principal plays A if $E[A|x(t)] > E[I|x(t)]$, which happens when $x(t) > x_N$. The principal thus plays α_N to make the agents indifferent between their two actions. The agents in turn best reply by mixing between C and H with probability x_N . The situation is identical each time the game is played and the result follows. $\square$

Proof of Theorem 3.2

Proof. Let the fraction of cheaters be denoted by $x(t)$ and denote the principal’s audit rate by $\alpha(t)$. The rates with which the C population and the audit rate evolve follow the values in Figure 1, and

are given by

$$\begin{aligned}\dot{\alpha}(t) &= \alpha(t) \left[- (c_1 x(t) + c_2 (1 - x(t))) - (c_3 (1 - \alpha(t)) x(t) + c_1 \alpha(t) x(t) + c_2 \alpha(t) (1 - x(t))) \right] \\ &= \alpha(t) (1 - \alpha(t)) [(c_3 + c_2 - c_1) x(t) - c_2]\end{aligned}\tag{18}$$

$$\begin{aligned}\dot{x}(t) &= x(t) [v_3 (1 - \alpha(t)) - (v_1 \alpha(t) (1 - x(t)) + v_2 (1 - \alpha(t)) (1 - x(t)) + v_3 (1 - \alpha(t)) x(t))] \\ &= x(t) (1 - x(t)) [(v_2 - v_3 - v_1) \alpha(t) + v_3 - v_2]\end{aligned}\tag{19}$$

Because of Assumptions 2.3 and 3.1, there are no equilibria on the boundary of the system described by (18) and (19). Instead, there is unique *interior* equilibrium which is obtained when $\dot{x}(t) = 0$ and $\dot{\alpha}(t) = 0$. At this equilibrium the pair (x, α) is equal to $(\frac{v_3 - v_2}{v_3 + v_1 - v_2}, \frac{c_2}{c_3 + c_2 - c_1})$, the same values for the Nash equilibrium of the repeated game in (1) (recall that there I have assumed that $v_1 = v_2$). The Hartman-Grobman linearization near the equilibrium gives the Jacobian of this system of equations evaluated at the equilibrium,

$$J = \begin{bmatrix} 0 & \frac{v_1(v_3 - v_2)}{(v_3 + v_1 - v_2)^2} (c_3 + c_2 - c_1) \\ \frac{c_2(c_3 - c_1)}{(c_3 + c_2 - c_1)^2} (v_2 - v_3 - v_1) & 0 \end{bmatrix}$$

Note that because of the structure and the relationship between the costs, particularly because $c_3 > c_1$ and $v_3 > v_2$, the entry at the top right is always positive while that on the bottom left is always negative. In particular, this system has a pair of pure imaginary eigenvalues, implying that the equilibrium is non-hyperbolic and is in fact a center of the dynamical system described by (18) and (19) (Perko (2001)).

Dropping the time argument in the following to reduce clutter, I can write the dependence of α on x as:

$$\frac{d\alpha}{dx} = \frac{\alpha(1 - \alpha) [(c_3 + c_2 - c_1)x - c_2]}{x(1 - x) [(v_2 - v_3 - v_1)\alpha + v_3 - v_2]}$$

This gives

$$\frac{((v_2 - v_3 - v_1)\alpha + v_3 - v_2)}{\alpha(1 - \alpha)} d\alpha - \frac{((c_3 + c_2 - c_1)x - c_2)}{x(1 - x)} dx = 0$$

Integrating and using $B(x, \alpha)$ to describe the solution to (18) and (19),

$$B(x, \alpha) = (v_2 - v_3 - v_1) \ln(1 - \alpha) + (v_3 - v_2) \ln \frac{\alpha}{1 - \alpha} - (c_3 + c_2 - c_1) \ln(1 - x) - c_2 \ln \frac{x}{1 - x}$$

Rearranging terms and exponentiating, I end up with

$$B(x, \alpha) = \alpha^{v_3 - v_2} (1 - \alpha)^{2(v_2 - v_3) - v_1} x^{-c_2} (1 - x)^{c_1 - c_3} \quad (20)$$

Now let (x, α) be a solution to the system (18) and (19). Then the rate of change of $B(x, \alpha)$ with respect to time is given by

$$\dot{B}(x, \alpha) = \dot{x} \frac{\partial}{\partial x} B(x, \alpha) + \dot{\alpha} \frac{\partial}{\partial \alpha} B(x, \alpha)$$

Claim .1. $\dot{B}(x, \alpha) = 0$ for any solution $(x(t), \alpha(t))$ to (18) and (19).

Proof. See below. □

Claim .1 implies that the orbits described by (20) are closed and correspond to constant levels of $B(x, \alpha)$, since their time derivative is zero. The Nash equilibrium of the game is a center of these orbits and of the dynamical system (18) and (19). □

Proof of Claim .1

Proof. I need to show that $\dot{B}(x, \alpha) = \dot{x} \frac{\partial}{\partial x} B(x, \alpha) + \dot{\alpha} \frac{\partial}{\partial \alpha} B(x, \alpha) = 0$, where

$$B(x, \alpha) = \alpha^{v_3 - v_2} (1 - \alpha)^{2(v_2 - v_3) - v_1} x^{-c_2} (1 - x)^{c_1 - c_3}$$

To reduce notation, I note that all the costs in the problem are constants that do not affect the derivatives with respect to x or α , I will employ the following shorthand notation. $a = c_2, b = c_3 - c_1, c = v_3 - v_2$, and $s = 2(v_2 - v_3)$. Substituting for these quantities and differentiating $B(x, \alpha)$

with respect to x , one gets

$$\frac{\partial}{\partial x}(\alpha^c(1-\alpha)^s x^a(1-x)^{-b}) = c(1-x)^{-b}x^{-a}(1-\alpha)^s \alpha^{c-1} - s(1-x)^{-b}x^{-a}(1-\alpha)^{s-1} \alpha^c \quad (21)$$

Similarly, the partial derivative $\frac{\partial}{\partial \alpha}B(x, \alpha)$ is given by

$$\frac{\partial}{\partial \alpha}(\alpha^c(1-\alpha)^s x^a(1-x)^{-b}) = (1-x)^{-b-1}x^{-a-1}(a(-1+x) + bx)(1-\alpha)^s \alpha^c \quad (22)$$

Using (18), (19), (21), and (22) in $\dot{B}(x, \alpha)$

$$\begin{aligned} \dot{B}(x, \alpha) &= (a - (a+b)x)(1-\alpha)\alpha \left(c(1-x)^{-b}x^{-a}(1-\alpha)^s \alpha^{-1+c} - s(1-x)^{-b}x^{-a}(1-\alpha)^{-1+s} \alpha^c \right) \\ &+ (1-x)x(c - (c+s)\alpha) \left(-a(1-x)^{-b}x^{-1-a}(1-\alpha)^s \alpha^c + b(1-x)^{-1-b}x^{-a}(1-\alpha)^s \alpha^c \right) \\ &= 0 \end{aligned}$$

□

Proof of Theorem 3.3

Proof. Consider a period of length T . We can rewrite (18) as

$$\begin{aligned} \frac{\dot{\alpha}(t)}{\alpha(t)(1-\alpha(t))} &= ((c_1 - c_3 - c_2)x(t) + c_2) \\ \int_0^T \frac{\dot{\alpha}(t)}{\alpha(t)(1-\alpha(t))} dt &= \int_0^T ((c_1 - c_3 - c_2)x(t) + c_2) dt \\ \ln \frac{\alpha(t)}{1-\alpha(t)} \Big|_{t=0}^{t=T} &= c_2 T - (c_1 - c_3 - c_2) \int_0^T x(t) dt \end{aligned}$$

Noting that $\alpha(0) = \alpha(T)$, we get

$$\begin{aligned} 0 &= c_2 T - (c_1 - c_3 - c_2) \int_0^T x(t) dt \\ \frac{c_2}{c_3 - c_2 - c_1} &= \frac{1}{T} \int_0^T x(t) dt \end{aligned}$$

Thus the average fraction of cheaters over any period is equal to that in (1). The same reasoning is used to show that

$$\frac{v_3 - v_2}{v_3 + v_2 - v_1} = \frac{1}{T} \int_0^T \alpha(t) dt,$$

which, given the assumption that $v_1 = v_2$, is again equal to the audit rate in (1). $\square$

Proof of Lemma 4.1

Proof. A bang-bang solution implies that $\alpha(t)$ takes on extremal values in its domain until the solution trajectory reaches a final state. I will 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

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

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

$$\alpha^*(t) = \begin{cases} 0, & \lambda(t) < \frac{c_2 + (c_1 - c_2 - c_3)x(t)}{(p + q)x(t)(1 - x(t))}; \\ 1, & \lambda(t) > \frac{c_2 + (c_1 - c_2 - c_3)x(t)}{(p + q)x(t)(1 - x(t))}; \\ [0, 1], & \lambda(t) = \frac{c_2 + (c_1 - c_2 - c_3)x(t)}{(p + q)x(t)(1 - x(t))}. \end{cases} \quad (23)$$

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 I show shortly, on the optimal control and state trajectories this case cannot happen except for precisely

a single pair (α^*, x^*) . □

Proof of Lemma 4.2

Proof. Denoting the function inside the integral in (10) 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,

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

After some algebra and simplifying the above, I get

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

which is a quadratic function in $x(t)$. Solving that equation and enforcing the constraint that $0 \leq x(t) \leq 1$, I 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 a function of the parameters of the problem. □

Proof of Theorem 4.3

Proof. I 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 statement of the theorem satisfies the necessary

conditions for an optimum. That an optimal policy exists follows from the boundedness of the cost per stage, $g(x(t), \alpha(t))$, and the continuity of the functions g and f on the compact sets $x(t)$ and $\alpha(t)$. The boundedness of the per-stage cost together with the presence of the discount factor r ensures that the value of the optimal solution is $< \infty$.

From Lemma 4.2, a necessary condition for the optimal path $x^*(t)$ to minimize (10) (and consequently, (4)), is that $x^*(t)$ is a constant, which I will denote by $\bar{x}$, where $\bar{x}$ is as given in the proof of Lemma 4.2. 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

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

Using Assumption 2.3, $\bar{x}$ cannot have a value on the boundary, and hence the only solution to the above equation is $\alpha^*(t) = \frac{q}{p+q}$. This implies, from (23), that $\lambda(t) = \frac{c_2 + (c_1 - c_2 - c_3)\bar{x}}{(p+q)\bar{x}(1-\bar{x})}$. 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 4.1, if $x^*(t) \neq \bar{x}$ and $\alpha(t) \neq 1$ then $\alpha(t) = 0$, in which case $\dot{x}(t) > 0$ and $x(t + \delta) > x(t)$ for δ small enough. Let $t + \delta = t_1 > t$, $x(t_1) > \bar{x}$ and $\alpha(t_1) = 0$, then for $t_2 > t_1$, $x(t_2) > x(t_1)$, i.e. the system moves farther from $\bar{x}$. However, because of Lemma 4.2, an optimal trajectory must 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 $x(t) < \bar{x}$. $\square$

Proof of Theorem 5.1

Proof. Denote the audit rate in the behavioral setting by α_B . From Theorem 4.3, α_B is given by $\frac{q}{p+q}$. Replacing p and q by their values from Section 2.2, I find that the audit rate α_B is given by

$$\alpha_B = \frac{\frac{v_3-v_1}{v_3}}{1 + \frac{v_3-v_1}{v_3}} \quad (24)$$

which is always *strictly less* than the Nash audit rate in (1). Now consider the situation as $r \rightarrow 0$. In the fully rational setting this does not affect the outcome, and the principal and the population keep playing the strategies prescribed by (1) forever. Under the behavioral setting however, letting $r \rightarrow 0$ has a drastic effect on the cutoff value $\bar{x}$. From Equation (11) in Lemma 4.2,

$$\lim_{r \rightarrow 0} \bar{x} = \frac{(c_2 - c_1)p - c_3q + (c_1 - c_2 - c_3)r + \sqrt{4c_2((c_2 - c_1)p - c_3q)r + ((c_1 - c_2)p + c_3q + (c_1r - c_2 - c_3)r)^2}}{2((c_2 - c_1)p - c_3q)}$$

Substituting for p and q , and taking the limit,

$$\lim_{r \rightarrow 0} \bar{x} = \frac{c_1 - c_2 + \frac{c_3(-v_2+v_3)}{v_3} - \sqrt{\left(c_1 - c_2 + \frac{c_3(-v_2+v_3)}{v_3}\right)^2}}{2\left(c_1 - c_2 + \frac{c_3(-v_2+v_3)}{v_3}\right)} = 0$$

which concludes the proof. □

Proof of Theorem 7.1

Proof. In an attempt to reduce notation, I will assume that the system evolves such that the share of a strategy that is performing better than average grows. Any quantitative derivation leads to the same qualitative results as long as agents switch to better strategies with some positive probability.

The system is then described by the following equations:

$$\dot{x}(t) = x(t)(1-x(t))((1-\alpha(t))a_2 - \alpha(t)a_1) \quad (25)$$

$$\dot{\alpha}(t) = \alpha(t)(1-\alpha(t))((1-x(t))b_1 - x(t)b_2) \quad (26)$$

Let $x(0) = \frac{b_1+\epsilon}{b_1+b_2}$, for any $\epsilon \in (0, b_2)$, then from (26),

$$\begin{aligned} \dot{\alpha}(t) &= \alpha(t)(1-\alpha(t)) \left(\frac{(b_2-\epsilon)b_1}{b_1+b_2} - \frac{(b_1+\epsilon)b_2}{b_1+b_2} \right) \\ &= -\epsilon(b_1+b_2)\alpha(t)(1-\alpha(t)) \\ &< 0. \end{aligned}$$

Similarly, one can show that $\dot{x}(t) > 0$ when $\alpha(0)$ fulfills the condition in the statement of the theorem. This implies that $x(\delta) > x(0)$ for any $\delta > 0$, and hence $\dot{\alpha}(t)$ at $t = \delta$ continues to be negative while $\dot{x}(t)$ continues to be greater than zero and $x(t)$ keeps increasing, leading to $\lim_{t \rightarrow \infty} (x(t), \alpha(t)) \rightarrow (1, 0)$. Thus the principal shifts almost all of the weight to the R strategy while more and more agents play action B , and the system converges to the equilibrium (B, R) as in the statement of the theorem. $\square$

Proof of Theorem 7.2

Proof. Using (26), one can write $\alpha(t)$ as

$$\alpha(t) = \frac{\dot{x}(t) - a_2x(t) + a_2x^2(t)}{(a_1 + a_2)(x(t) + x^2(t) - x(t))}$$

The principal's problem becomes

$$\min_{x(t)} \int_0^\infty e^{-rt} \left(\frac{a_2b_1(1-x(t)) + a_1b_2x(t) + \frac{(b_1-(b_1+b_2)x(t))\dot{x}(t)}{(1-x(t))x(t)}}{a_1 + a_2} \right) dt \quad (27)$$

where $r > 0$ is a discount factor like before. Using the Euler-Lagrange equation to solve (27) leads to the following condition that should be satisfied by the optimal $x(t)$

$$\frac{e^{-rt} (-b_1 r + (a_1 b_2 e^{rt} - a_2 b_1) x(t))}{(a_1 + a_2) x(t)} = 0 \quad (28)$$

Denoting the optimal $x(t)$ by $x^*(t)$, we get

$$x^*(t) = \frac{b_1 r}{a_1 b_2 e^{rt} - a_2 b_1} \quad (29)$$

Then,

$$\lim_{r \rightarrow 0} x^*(t) = \lim_{r \rightarrow 0} \frac{b_1 r}{a_1 b_2 e^{rt} - a_2 b_1} = 0 \quad (30)$$

Thus as the discount factor gets higher (by having r approach zero), the system converges to a state where no one plays B ,¹¹ and the principal's optimal action is to set $\alpha(t) = 1$ indefinitely. $\square$

Proof of Proposition 8.1

Proof. The proof follows Section 4.2.1. Let the threshold in the statement of that theorem be given by x_N , and let the fraction of the behavioral population playing C be x . When $x + \rho > x_N$, the principal's payoff is maximized by setting $\alpha = 1$. Thus $p_c^* = 0$ in this case, since no rational player would want to cheat as they will get audited with certainty. Similarly, when $\rho < x_N - x$, the principal's optimal action is given by (5), and $\alpha = 0$, leading to $p_c^* = 1$. Finally, when the value of p_c^* decides whether $x + p_c^* \rho > x_N$ or $x + p_c^* \rho < x_N$, i.e. the situation is such that the fraction of cheaters in the rational population determines which side of x_N the total number of C players falls on, the rational agents maximize their utility by playing C with a probability that makes the principal indifferent between auditing and ignoring: If $p_c^* > \frac{x_N - x}{\rho}$, then the principal's optimal action is $\alpha = 1$. Conversely, if $p_c^* < \frac{x_N - x}{\rho}$ then the principal sets $\alpha = 0$, and the agents could increase p_c^* slightly and do better. It follows that the equilibrium value for p_c^* is the one that makes $E[C|x + p_c^* \rho, \alpha] = E[H|x + p_c^* \rho, \alpha]$, which happens exactly when p_c^* is as in the statement of the

¹¹Of course, it would require infinite time for $x(t)$ to reach zero.

theorem, and the principal then plays the audit rate α_N that makes $E[A|x, p_c^*] = E[I|x, p_c^*]$. $\square$

Proof of Proposition 8.2

Proof. It is sufficient to show that there exists a solution –not necessarily optimal– where $\alpha(t) < \alpha_N$ and $x(t) + p_c^* \rho < x_N$. One such solution exists when the whole rational population plays C , i.e. $p_c^* = 1$. The equation of motion of the behavioral population becomes

$$\dot{x}(t) = x(t)(v_3(1 - \alpha(t)) - ((x(t) + \rho)v_3(1 - \alpha(t)) + (1 - x(t))v_1)) \quad (31)$$

Replacing $x(t)$ by $x_N - \rho$, the value of $\alpha(t)$ that stabilizes the *total population* of C players at x_N , denoted by $\bar{\alpha}$, can be found from (31), where

$$\bar{\alpha} = \frac{v_3 - v_1 + v_1(x_N - \rho) - v_3(x_N - \rho) - v_3\rho}{v_3(1 - x_N)} \quad (32)$$

Obviously, $\bar{\alpha} < \alpha_N$ or else $p_c^* < 1$ by Lemma 8.1, in which case the construction above fails. To verify that it is indeed the case, replace x_N by its value from (1) and subtract the Nash audit rate α_N from (32), leading to

$$\bar{\alpha} - \alpha_N = \frac{(c_3 + c_2 - c_1)v_1\rho}{(c_1 - c_3)v_3} \quad (33)$$

which is less than zero as $c_3 > c_1$. Therefore, there exists a solution that gives the same fraction of cheaters as the Nash solution but audits at a value that is less than the Nash rate. $\square$

Proof of Proposition 8.3

Proof. This follows immediately from Lemma 8.1. Assume that at some point in time $\alpha^*(t) < \alpha_N$, then $p_c^*(\alpha(t)) = 1$. Of course, if $x(t) > 0$, $x(t) + \rho$ increases, and $\alpha^*(t)$ increases. Even if the principal drives $x(t)$ to zero, he would still be playing a fully rational game against a fraction ρ of the population, and any $\alpha(t) < \alpha_N$ is suboptimal by Theorem 2.2. It follows that $\alpha^*(t) \geq \alpha_N$ and the rational population plays as in Lemma 8.1. $\square$

Proof of Lemma 4.2

Proof. Denoting the function inside the integral in (10) 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,

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

After some algebra and simplifying the above, I get

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

which is a quadratic function in $x(t)$. Solving that equation and enforcing the constraint that

$0 \leq x(t) \leq 1$, I 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 a function of the parameters of the problem. □